UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA
Dipartimento di Ingegneria dell'Informazione
Corso di Laurea in Ingegneria Informatica e dell'Automazione

TESI DI LAUREA

Progettazione e implementazione di una campagna di data science sui dati di vendita di un'importante azienda di gioielli

Design and implementation of a data science campaign on sales data for a major jewelry company

Relatore

Prof. Domenico Ursino

Candidato

Federica Paganica

Correlatore

Prof. Francesco Cauteruccio

ANNO ACCADEMICO 2022-2023

Ai miei genitori,
che hanno sempre creduto in me

Sommario

In un mondo sempre più digitalizzato e interconnesso, la Data Analytics riveste un ruolo di fondamentale importanza. Grazie ai grandi volumi di dati di cui si dispone oggi, quest'ultima consente di estrarre informazioni preziose che possono influenzare non solo le decisioni aziendali, ma anche le strategie per il futuro. Questo elaborato mira a sottolineare l'utilità e l'efficacia della Data Analytics, dimostrando come l'analisi accurata di grandi quantità di dati possa fornire una panoramica dettagliata dei fenomeni aziendali e offrire ulteriori linee guida per la pianificazione strategica. In questa tesi, abbiamo condotto una campagna di Data Science sulle vendite dell'azienda marchigiana Bros Manifatture. Dopo aver descritto i dati, abbiamo eseguito operazioni di ETL, utilizzando il software Power BI, per pulire gli stessi, organizzarli e predisporli alle analisi successive. Mediante l'uso di tecniche di Data Visualization, abbiamo realizzato report interattivi sulle vendite ai clienti e al consumatore finale. Questi strumenti hanno messo in risalto le tendenze e i pattern aziendali, offrendo un quadro dettagliato e comprensibile che potrà rappresentare la base per future strategie di business.

Keyword: Data Analytics; Big Data; Data Science; Business Intelligence; Extract, Trasform and Load; Power BI; Data Visualization.

Indice

Elenco delle figure

Introduzione

> *"Scegliere come interpretare i dati è molto simile a fare un ritratto. Puoi dipingere con i numeri come Monet con i colori. Ci sono sempre opportunità di capire il mondo in modo nuovo, di connetterti con esso e di trarne un beneficio."*
>
> (Susan Wojcicki, CEO di YouTube.)

Al giorno d'oggi, l'analisi dei Big Data si è imposta come uno strumento indispensabile nel mondo delle aziende, permettendo di attingere a un vasto panorama di informazioni per prendere decisioni informate e strategiche. Questo processo richiede un tocco artistico: come sottolinea Susan Wojcicki, CEO di YouTube, l'interpretazione dei dati può essere paragonata alla pittura di un ritratto. Non si tratta semplicemente di manipolare numeri, ma di utilizzarli per creare un'immagine che rivela qualcosa di unico e importante.

La tesi che segue esplora il ruolo della Big Data Analytics nella conduzione di una campagna di Data Science, sottolineando come, similmente a Monet, possiamo usare i numeri non solo per rappresentare la realtà, ma anche per aprirci nuove prospettive su di essa e trarne vantaggio.

L'era digitale ha portato ad un'enorme quantità di dati memorizzati in tutto il mondo. Questo grande insieme di informazioni è stato definito come "Big Data". In questo contesto, la tecnologia ha sviluppato nuove tecniche e metodologie analitiche che sono state create per gestire questa ingente quantità di dati. Infatti, la disponibilità di grandi quantità di dati richiede un'analisi approfondita e mirata; altrimenti i dati rimarranno inutilizzati. Le soluzioni presenti sul mercato possono gestire il volume dei dati e sfruttare al meglio le informazioni raccolte.

L'utilizzo dei Big Data è associato in modo massivo a molte attività e processi aziendali. In altre parole, i dati sono un elemento fondamentale per il funzionamento quotidiano delle aziende, poiché consentono di analizzare e valutare informazioni cruciali per la loro gestione e la loro crescita. Il termine "utilizzo massivo" suggerisce che le aziende stanno raccogliendo, analizzando e interpretando grandi quantità di dati per migliorare i propri processi. Ciò riguarda molteplici ambiti, come, ad esempio, il marketing, la finanza, la gestione delle risorse umane e la produzione.

Inoltre, la crescente importanza dei dati ha portato all'emergere di nuove figure aziendali specializzate nella gestione di questi ultimi, come i Data Analyst e i Data Scientist. Queste figure sono impegnate nella raccolta e nel trattamento dei dati, nella loro analisi e nell'elaborazione di report e previsioni basate sui di essi.

In definitiva, l'utilizzo delle tecniche di analisi dei Big Data è cruciale nell'ambito di un mondo sempre più digitale e connesso. La loro applicazione mirata può aprire nuo-

ve prospettive di business, migliorare i processi decisionali e ottimizzare le strategie di investimento.

Alla luce di tutto ciò, in questa tesi, proponiamo una campagna di analisi dei Big Data sulle vendite dell'azienda marchigiana Bros Manifatture. Dopo un dettagliato percorso di acquisizione dei dati forniti dall'azienda, abbiamo condotto operazioni di ETL (Extract, Transform, Load) al fine di pulire, modellare e preparare i dati per le successive fasi di analisi. Tutti questi processi sono stati gestiti con l'ausilio del software di Business Intelligence Power BI.

Successivamente, utilizzando tecniche di Data Visualization, sono stati realizzati vari report interattivi sulle vendite ai clienti e su quelle al consumatore finale. Questi strumenti hanno fornito una panoramica dettagliata e facilmente interpretabile delle prestazioni di vendita, evidenziando le tendenze principali e fornendo una base solida per future strategie di business.

In conclusione, questo lavoro sottolinea la crucialità della Data Analytics nel moderno panorama aziendale. Ogni passaggio svolto durante il lavoro di tesi ha dimostrato come una gestione efficace dei dati possa portare a una migliore comprensione delle dinamiche aziendali e a una pianificazione strategica più informata e mirata.

La presente tesi è composta da cinque capitoli strutturati come di seguito specificato:

- Il Capitolo 1 offre un'esaustiva panoramica sul tema della Data Analytics. Esso fornisce una comprensione dei Big Data presentando la teoria delle 3V, il modello delle 5V e la differenza tra Data Analytics e Data Analysis. Infine, si discute del ciclo di vita della Big Data Analytics analizzando le sue 9 fasi.

- Il Capitolo 2 si concentra sul software di Business Intelligence Power BI analizzando la sua architettura e l'importanza di Power Query nel processo di data cleaning. Si esplora, anche, l'aspetto della Data Visualization, i diversi tipi di visualizzazioni, i filtri e le misure DAX.

- Nel Capitolo 3 si esplorano i vari tipi di dati: strutturati, non strutturati, semistrutturati e metadati. Esso evidenzia il loro ruolo chiave nell'analisi delle informazioni. Inoltre, si analizza il processo di Estrazione, Trasformazione e Caricamento (ETL) e si presenta un dettagliato resoconto delle operazioni di ETL condotte sui dataset forniti da Bros Manifatture.

- Nel Capitolo 4 si accenna alle varie categorie di Data Analytics, con un focus sull'analisi descrittiva. Si descrivono le analisi effettuate sui dataset aziendali di Bros Manifatture attraverso report realizzati in Power BI, con particolare attenzione allo studio del cliente e del consumatore finale.

- Nel Capitolo 5 si approfondisce la tematica della Business Intelligence, analizzando il suo impatto sul panorama aziendale moderno, giungendo a una riflessione sugli aspetti positivi e negativi del percorso di Big Data Analytics intrapreso, in modo da offrire una panoramica completa del lavoro svolto.

CAPITOLO 1

Introduzione alla Big Data Analytics

In questo capitolo si intende offrire un'esaustiva panoramica sul tema della Data Analytics. Inizieremo fornendo un quadro generale riguardo ai Big Data, partendo dalle loro caratteristiche essenziali per poi approfondire analizzando la teoria delle 3V, proposta da Doug Laney, per giungere, quindi, al modello delle 5V. Successivamente, ci concentreremo sulla definizione di Data Analytics, mettendo in evidenza le peculiarità che la distinguono dalla Data Analysis. Infine, ci dedicheremo allo studio dell'intero ciclo di vita della Big Data Analytics, analizzando in dettaglio ciascuna delle 9 fasi che lo compongono.

1.1 Cosa sono i Big Data?

Negli ultimi anni l'importanza dei dati è cresciuta in maniera esponenziale a tal punto da diventare una vera e propria risorsa per molti settori produttivi. Grazie ai progressi raggiunti in ambito informatico e tecnologico, le aziende, utilizzando metodologie di Big Data Analytics, raccolgono quanti più dati possibili al fine di elaborarli e sfruttarli al loro vantaggio. Questo processo risulta fondamentale per perseguire obiettivi come:

- incremento delle vendite;
- ottimizzazione dei processi decisionali;
- miglioramento dell'engagement con il cliente;
- scoperta di nuovi mercati;
- previsioni accurate.

Da sottolineare è la crescita dei dati che, in quest'ultimo decennio, è stata davvero esponenziale: nel 2018 la mole totale di dati generati nel mondo è stata di 29 ZB con un aumento di oltre 10 volte superiore rispetto al 2011, ma ancora più sorprendenti sono le previsioni per il 2025, quando il volume complessivo dei dati raggiungerà la soglia dei 175 ZB (Figura 1.1).

Il termine "Big Data" racchiude un significato molto complesso e viene utilizzato in maniera molto ampia; esso è associato in modo massivo a molte attività e processi aziendali. Per comprendere meglio l'argomento ci possiamo soffermare sulla seguente definizione:

> *"I Big Data rappresentato una raccolta di dati così estesa in termini di volume, velocità e varietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore."* *De Mauro* et al. *[2016]*

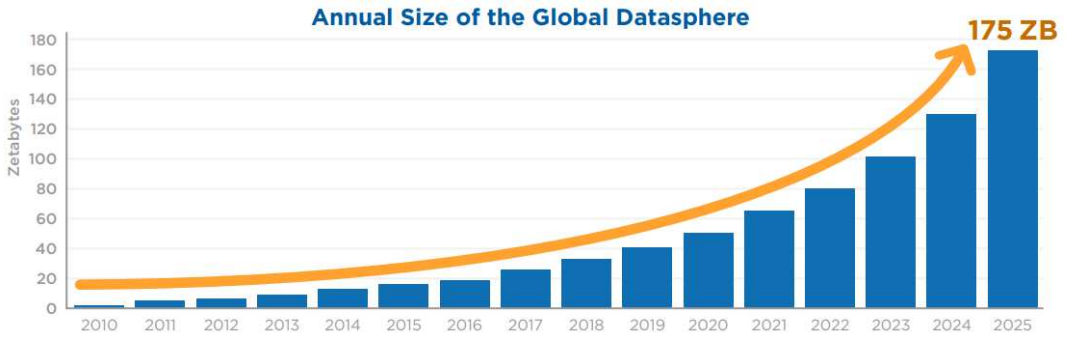

Figura 1.1: Annual Size of the Global Datasphere

Dalle parole di questa definizione possiamo cogliere quattro caratteristiche essenziali del concetto di Big Data (Figura 1.2):

- *dati*, poiché i Big Data sono una mole di informazioni da cui possiamo estrapolare i contenuti utili;
- *tecnologie*, in quanto vi è la necessità dell'uso di strumenti tecnologici adeguati per poter utilizzare i dati raccolti;
- *metodi*, poiché è indispensabile l'uso di determinati metodi analitici per l'analisi dei dati;
- *impatto*, in quanto l'utilizzo dei Big Data, consentendo di identificare trend e opportunità di mercato, può portare alla creazione di valore per le aziende.

Il termine "Big Data" si riferisce a una grande quantità di dati digitali, strutturati e non strutturati, che provengono da fonti diverse, come, ad esempio, i social network, le transazioni bancarie, gli acquisti online, i sensori e le analisi geospaziali.

Oggigiorno viviamo in un'epoca in cui circola una grandissima mole di dati digitali che provengono da fonti diverse. Alcuni esempi nella nostra vita quotidiana sono: i social network come Instagram, Youtube, WhatsApp, dove ogni giorno miliardi di persone creano dati tramite post, foto, reel. Tuttavia, non possiamo tralasciare i dati generati da macchine "intelligenti" grazie alla recente diffusione dell'IoT (Internet Of Things). Quest'ultimo lo ritroviamo:

- Nelle nostre case dove, grazie ai moderni sistemi di automazione domestica, attraverso l'utilizzo di interruttori e sensori, è possibile monitorare e azionare, per esempio, l'illuminazione, la climatizzazione, i sistemi di sicurezza, e gli elettrodomestici anche da remoto.
- Nelle industrie, dove tale tecnologia è presente in molti esempi di produzione intelligente dell'Industria 4.0 (la quarta rivoluzione industriale) e rende il processo produttivo più veloce, aumentandone l'efficienza e la sostenibilità.
- Nelle smart city, dove questi dispositivi, impiegati, ad esempio, nei trasporti pubblici, nella distribuzione dell'energia, nella sicurezza urbana e nel monitoraggio ambientale, sono in grado di garantire un'elevata qualità di vita ai cittadini.

La quantità di dati generati dalle persone e dalle macchine intelligenti è così smisurata da condurci alla definizione di Big Data.

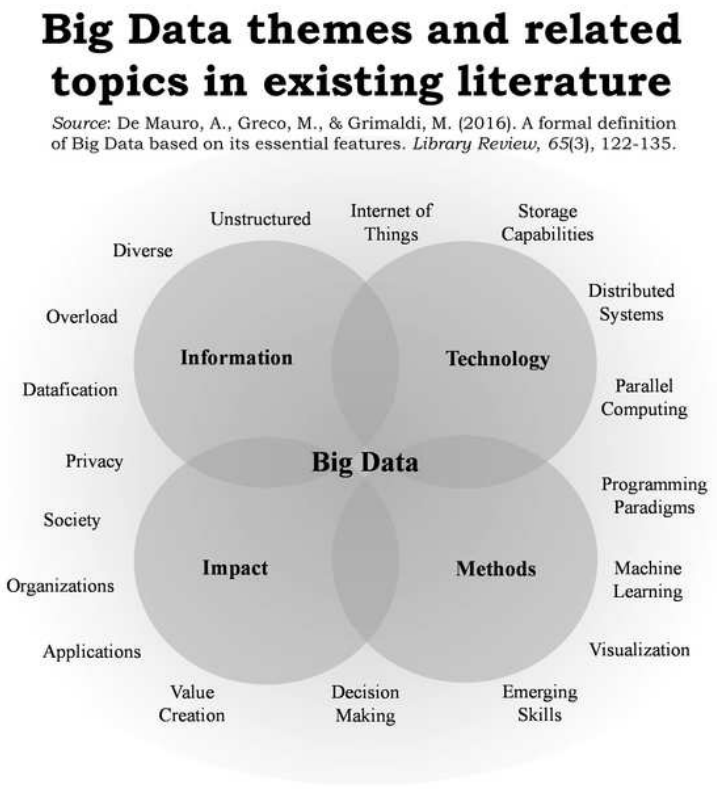

Figura 1.2: Temi relativi ai Big Data e argomenti correlati

1.2 Le 5 V

Il fenomeno dei "Big Data", ossia la raccolta, l'accumulo e l'analisi di ingenti quantità di dati, presenta alcune caratteristiche frequenti. Queste ultime possono essere riassunte nelle 3V, il Volume, la Varietà e la Velocità, identificate inizialmente nel 2001 da Doug Laney che le descrisse in un report dell'azienda Meta Group.

A queste 3V si sono aggiunte, successivamente, altre 2V, ovvero la Veracità e il Valore (Figura 1.3).

1.2.1 Volume

Il volume dei Big Data si riferisce all'enorme quantità di informazioni che vengono generate e raccolte. Molteplici sono i dispositivi che creano dati in continuazione a cui possiamo sommare quelli generati dagli utenti online attraverso i social network, dalle attività bancarie, dalle registrazioni mediche e da tutti i settori della Pubblica Amministrazione. Nel mondo attuale buona parte delle attività economiche si sono trasferite su Internet; infatti un'altra fonte rilevante di dati sono le transazioni effettuate online dai consumatori tramite gli acquisti sui siti di e-commerce (nella Figura 1.4 viene fornito un grafico che rappresenta le vendite al dettaglio online, in tutto il mondo, a partire dall'anno 2014 fino al 2023). La mole di informazioni può arrivare a superare i petabyte o gli exabyte e richiede specifiche architetture di memorizzazione per gestire i dati in modo efficiente.

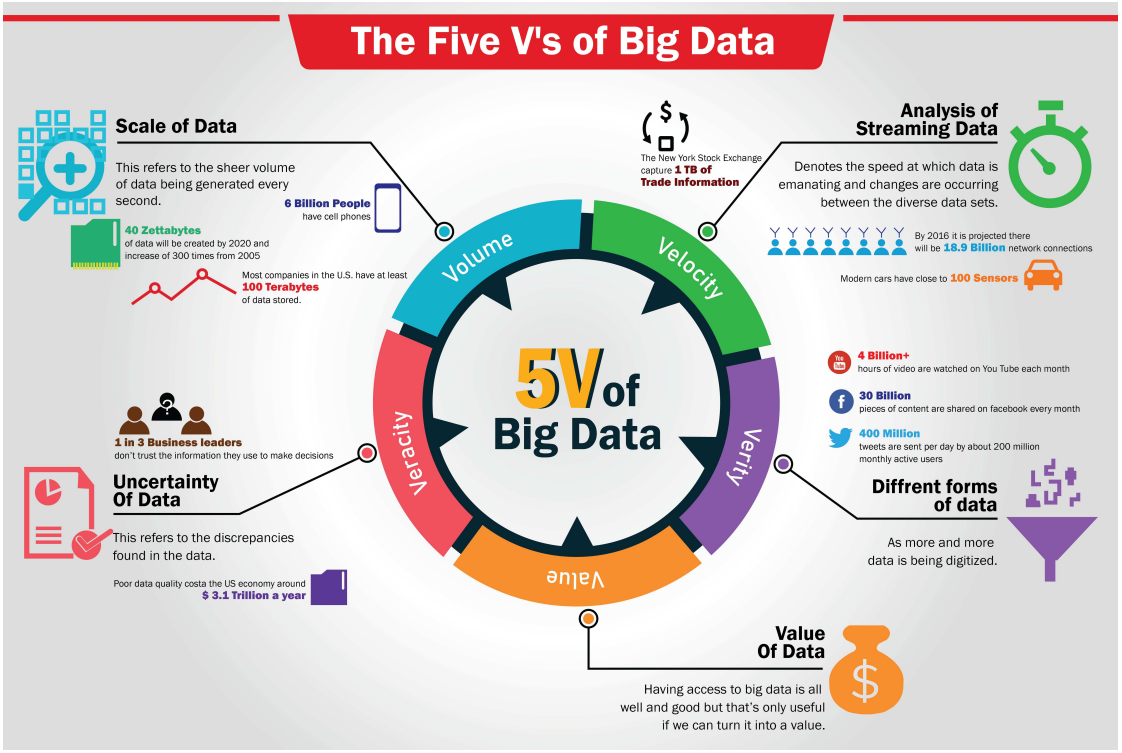

Figura 1.3: Le 5V dei Big Data

1.2.2 Varietà

Un'ulteriore caratteristica dei Big Data è la varietà del loro formato. Basti pensare ai dati che vengono generati continuamente dalle persone attraverso post, foto e video sui social media, totalmente diversi dalla tabelle numeriche che ritroviamo in un tradizionale database. Le numerose tipologie di dati comprendono dati strutturati, come quelli relativi alle transazioni finanziarie o alle informazioni mediche, dati semi-strutturati in forma di e-mail, e dati non strutturati, come i tweet o le conversazioni in una chat.

1.2.3 Velocità

La velocità dei Big Data si riferisce alla rapidità con cui vengono generati e trasmessi i dati, che vanno ad accumularsi in dataset di enormi dimensioni in brevissimo tempo. Ad esempio, un'automobile intelligente, come la Tesla, può produrre circa 80 Gb di dati in un solo minuto. Le informazioni devono essere processate in tempo reale per avere un impatto immediato sulla vita degli individui o sul business delle aziende. In particolare, un'azienda, affinché il flusso di dati venga processato in un tempo ragionevole, deve adottare soluzioni di data processing flessibili ed una notevole capacità di memorizzazione delle informazioni.

1.2.4 Veracità

La Veracità dei Big Data si riferisce alla loro qualità e affidabilità. È essenziale che i dati siano veri affinché risultino utili per prendere decisioni importanti. I dataset, prima del loro utilizzo, devono essere certificati per la qualità in modo da eliminare dati non idonei e rimuovere il rumore. È possibile distinguere dati che fanno parte del segnale (contengono informazioni significative e hanno un alto valore) e dati che fanno parte del rumore (non possono essere convertiti in informazioni e, quindi, non hanno alcun valore). I dati che

Figura 1.4: Retail E-Commerce Sales Worldwide 2014 – 2023. Source: Statista

presentano un alto rapporto segnale-rumore sono più affidabili. La veracità dei dati può essere garantita attraverso tecniche di validazione e di controllo della qualità.

1.2.5 Valore

Il valore dei Big Data consiste nel loro potenziale di generare informazioni utili per le aziende e le organizzazioni. I dati vengono raccolti ed elaborati al fine di creare informazioni che possono essere utilizzate per migliorare la produttività, ridurre i costi, ottimizzare i processi e guadagnare nuove opportunità di business. Questo attributo è strettamente correlato alla veracità: maggiore è l'affidabilità dei dati e maggiore sarà il loro valore per l'azienda. Tuttavia il valore non è solo legato alla fedeltà ma anche alla tempestività: i dati hanno una "data di scadenza" e, quindi, il loro valore decresce nel tempo. Ad esempio, una quotazione finanziaria aggiornata ogni 20 minuti è praticamente inutile per chi deve compiere operazioni in borsa, dato che le fluttuazioni del mercato avranno già influenzato il prezzo. Al contrario, una quotazione aggiornata in millisecondi permette di prendere decisioni tempestive e puntuali. È importante, perciò, garantire la rapidità di elaborazione dei dati, in quanto il tempo è inversamente correlato al valore: più tempo si impiega nel processare i dati, meno essi saranno utili per l'azienda. Inoltre, i dati obsoleti o non aggiornati possono portare a decisioni sbagliate, rallentando il processo di decision-making e compromettendo la qualità dei risultati.

1.3 La Data Analytics

L'era digitale, con l'avvento dei Big Data, ha determinato la necessità di sviluppare nuove tecniche e metodologie analitiche, create per poter gestire quest'enorme volume di dati. La disponibilità di grandi quantità di dati richiede un'analisi approfondita e mirata; ciò consente di rilevare trend e pattern che, altrimenti, non sarebbero visibili, migliorando, così, i processi decisionali e riducendo i rischi di errori. Con l'espressione "Data Analytics" si intende la disciplina che racchiude il ciclo di vita completo dei dati che si può riassumere in raccolta, pulizia, organizzazione, memorizzazione e analisi. Inoltre, questa disciplina comprende lo sviluppo di metodi di analisi, tecniche scientifiche e tool automatici.

Le tecniche di analisi possono essere applicate a una vasta gamma di settori, da quello sanitario a quello finanziario, dal commercio ai social network. Ad esempio, nell'ambito sanitario, è possibile utilizzare la Data Analytics per individuare i sintomi comuni tra i pazienti, permettendo una diagnosi più precisa e un trattamento più efficace. Allo stesso modo, le analisi di dati finanziari possono aiutare le aziende a valutare le prospettive a lungo termine e a individuare eventuali rischi finanziari futuri.

1.4 La differenza tra Data Analytics e Data Analysis

L'utilizzo dei Big Data è in forte espansione e i dati sono ormai diventati una preziosa risorsa nel mondo del business e in molte altre sfere della vita. A causa della crescente popolarità di questo fenomeno, i termini "Data Analytics" e "Data Analysis" vengono spesso utilizzati in modo intercambiabile. Ma esiste una vera differenza tra i due?

La Data Analytics è una disciplina molto ampia che comprende tutti i processi riguardanti il ciclo di vita dei dati. La Data Analysis, invece, si concentra sul processo di ispezione dei dati allo scopo di trasformarli in informazioni significative e utili nel processo decisionale: l'obiettivo è trovare trend, relazioni, pattern e previsioni future. La Data Analysis è, quindi, un singolo aspetto ma molto importante della Data Analytics.

1.5 Il ciclo di vita della Big Data Analytics

La Big Data Analytics e l'analisi dei dati tradizionale sono diverse tra loro perchè sono diversi il volume, la velocità e la varietà delle informazioni che devono essere elaborate. È fondamentale, quindi, utilizzare una metodologia ben strutturata, procedendo per gradi nell'organizzazione delle attività. Il ciclo di vita della Big Data analytics si suddivide in varie fasi, esattamente nove (Figura 1.5), che analizzeremo, di seguito, in modo approfondito.

1.5.1 Business Case Evaluation

Il ciclo di vita della Big Data Analytics ha origine da un preciso business case dal quale dobbiamo estrapolare le motivazioni e gli obiettivi dell'analisi. Per iniziare, bisogna creare, validare e approvare il business case. Questo stadio è cruciale per i decision maker che possono individuare le risorse di business e le difficoltà che dovranno essere affrontate prima di procedere con i task di analisi effettivi.

Durante la valutazione del business case è possibile identificare i KPI (Key Performance Indicator), ovvero gli indicatori utilizzati per valutare il successo di un'attività. Tuttavia, se i KPI non sono subito disponibili, è necessario rendere SMART gli obiettivi del progetto di analisi; in questo contesto, il termine "SMART" sta per Specifici, Misurabili, Raggiungibili, Rilevanti e Tempestivi.

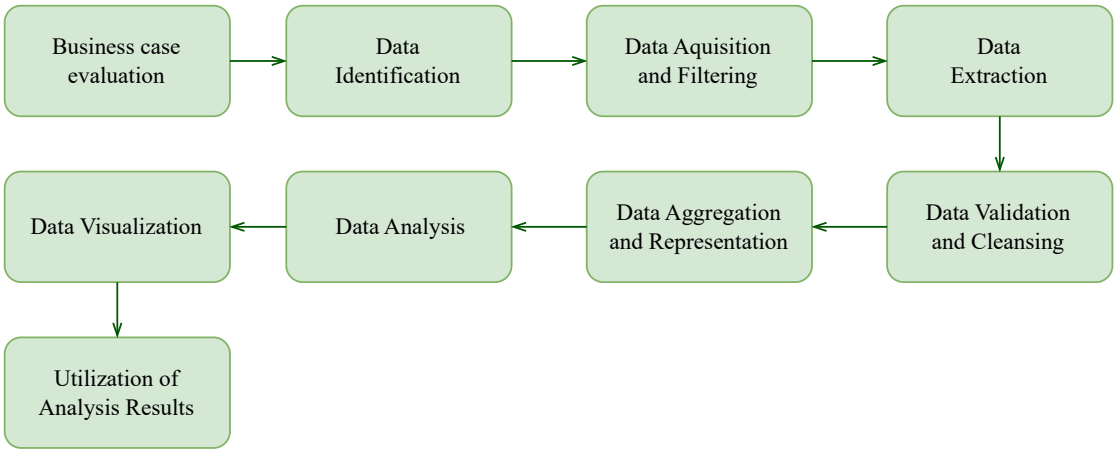

Figura 1.5: Il ciclo di vita della Big Data Analytics

Inoltre, in questa fase, si stabilisce se si è difronte ad un problema di Big Data o meno; per comprenderlo è necessario considerare le cinque caratteristiche dei Big Data: Volume, Varietà, Velocità, Veracità e Valore.

Un altro degli obiettivi di questo stadio è determinare il budget necessario per il progetto, che deve comprendere tutte le spese per l'acquisto di strumenti, hardware e formazione del personale. È importante stabilire in anticipo ogni acquisto richiesto, in modo tale da avere una visione chiara del budget totale e confrontarlo con i benefici attesi.

1.5.2 Data Identification

La fase di Data Identification è finalizzata a identificare i dataset richiesti per un progetto di analisi e individuare una varietà di sorgenti di dati, in modo da aumentare la possibilità di trovare pattern e correlazioni nascoste.

I dataset possono provenire da fonti interne ed esterne all'impresa. Nel caso di dataset interni, viene compilata una lista di dataset disponibili dalle sorgenti interne, tra cui i data mart e i sistemi operativi; questa lista viene, poi, confrontata con una specifica di dataset predefinita. Nel caso di dataset esterni, viene compilata una lista di possibili data provider di terze parti, tra cui data market e dataset disponibili al pubblico.

1.5.3 Data Aquisition and Filtering

Nella fase di Data Aquisition and Filtering vengono raccolti i dati dalle varie sorgenti precedentemente identificate. Tali dati vengono, poi, filtrati in modo da eliminare quelli senza valore o corrotti. Per dati "corrotti" si intendono record con valori nulli, privi di senso o non validi.

Prima di filtrare i dati, è necessario effettuare un backup del dataset originale poiché potrebbe essere utile per ulteriori analisi future.

Inoltre, i dati acquisiti devono essere sempre resi persistenti prima o dopo l'analisi, in base al tipo di analitica utilizzata (batch o realtime).

Questo processo può essere migliorato attraverso l'aggiunta di metadati, sia per le sorgenti di dati interne che per quelle esterne. I metadati devono essere scritti in un formato comprensibile dalla macchina e devono aver superato i vari stadi di analisi.

1.5.4 Data Extraction

È frequente che una porzione delle informazioni si presenti in un formato non conforme alle soluzioni di Big Data; il rischio di dover gestire diversi tipi di dati eterogenei cresce, in particolar modo quando si tratta di fonti esterne. L'obiettivo di questa fase è di prelevare informazioni di diversa natura e convertirle in un formato adatto all'analisi dei dati. Ovviamente, i processi di estrazione e conversione variano in base al tipo di analisi e alle caratteristiche della soluzione di Big Data impiegata.

1.5.5 Data Validation and Cleansing

È fondamentale rilevare che i dati non validi possono compromettere gravemente la qualità delle conclusioni tratte dalle analisi effettuate.

Nel contesto delle informazioni aziendali tradizionali, si registrano una struttura prestabilita dei dati e un processo di pre-convalida. Tuttavia, nel panorama dei Big Data, si assiste spesso all'acquisizione di dati non strutturati, i quali possono essere privi di criteri di validità.

La complessità intrinseca dei dati potrebbe, per di più, rendere arduo definire un insieme di vincoli di validazione appropriati. In tale fase, risulta cruciale l'attuazione del processo di Data Validation and Cleansing, il quale mira a consolidare regole stringenti per la validazione e a eliminare eventuali dati riscontrati come non validi.

Le soluzioni basate sui Big Data sono spesso caratterizzate dall'acquisizione di dati ridondanti provenienti da diversi dataset. Tale ridondanza può essere impiegata strategicamente per studiare le connessioni esistenti tra i vari dataset, al fine di elaborare parametri di validazione accurati e di fornire completamento ai dati validi che risultano mancanti.

Per quanto riguarda l'analitica batch, il processo di convalida e pulizia dei dati può essere realizzato attraverso un'operazione di ETL offline. Invece, nel caso di analitica real-time, è necessario un sistema in-memory più avanzato che possa verificare e correggere i dati man mano che giungono dalla fonte. L'origine dei dati può avere un impatto significativo sulla determinazione dell'accuratezza e della qualità delle informazioni incerte. Tuttavia, i dati che inizialmente risultano non validi potrebbero rivelarsi utili in quanto indicatori di pattern e tendenze nascoste.

1.5.6 Data Aggregation and Representation

I dati possono essere distribuiti su diversi dataset; ciò richiede che questi ultimi siano collegati attraverso elementi comuni, come la data o l'identificatore. In alcune situazioni, gli stessi attributi possono comparire in più insiemi di dati. È fondamentale disporre di un sistema di riconciliazione dei dati o stabilire quale dataset rappresenta il valore corretto. La fase di Data Aggregation and Representation si focalizza sull'integrazione di diversi dataset per ottenere una visione unificata.

Questa attività può diventare complessa a causa delle differenze relative alla struttura dei dati (poiché il formato può essere identico, ma il modello dei dati diverso), e alla semantica, dove un valore etichettato diversamente in due insiemi di dati potrebbe avere lo stesso significato (ad esempio, "surname" e "lastname").

Le grandi quantità di dati elaborate dalle soluzioni di Big Data possono rendere l'aggregazione dei dati un'operazione onerosa in termini di tempo e risorse. Riconciliare queste differenze può richiedere una logica complessa eseguita automaticamente senza la necessità di intervento umano. Durante questa fase, è importante considerare le future richieste di analisi dei dati per promuovere la riutilizzabilità delle informazioni.

Sia nel caso in cui l'aggregazione dei dati sia richiesta, sia nel caso in cui non lo sia, è fondamentale comprendere che le informazioni possono essere archiviate in diverse forme. In base al tipo di analisi, una forma potrebbe essere più adatta dell'altra.

1.5.7 Data Analysis

La fase di analisi dei dati è focalizzata sull'elaborazione e l'interpretazione dei dati raccolti. Durante questa fase, il processo può essere intrinsecamente iterativo, soprattutto se l'analisi è di natura esplorativa. In tal caso, l'analisi viene effettuata ripetutamente fino a quando non si identificano schemi o correlazioni pertinenti.

A seconda degli obiettivi dell'analisi, questa fase può variare in termini di complessità. In alcuni contesti, può essere sufficiente interrogare un insieme di dati per aggregare le informazioni e confrontarle con altre. Tuttavia, in situazioni più complesse, potrebbe essere necessario utilizzare tecniche avanzate di data mining e analisi statistica per identificare schemi, anomalie o generare modelli statistici o matematici che delineano le relazioni tra le variabili.

L'analisi dei dati può essere classificata in due categorie principali: confermativa o esplorativa, quest'ultima strettamente legata al data mining.

L'analisi confermativa si basa su un approccio deduttivo, in cui si propone una causa iniziale per il fenomeno in esame. Questa causa proposta, o supposizione, viene definita ipotesi. Successivamente, i dati vengono analizzati per verificare o confutare l'ipotesi e fornire risposte definitive a domande precise. In questo caso, si utilizzano soltanto tecniche di campionamento dei dati, ignorando risultati inaspettati o anomalie.

Al contrario, l'analisi esplorativa dei dati è un approccio induttivo correlato al data mining, in cui non si effettuano ipotesi o assunzioni preliminari. I dati vengono esaminati per sviluppare una comprensione della causa del fenomeno in questione. Nonostante questo metodo potrebbe non fornire risposte definitive, esso offre comunque una direzione generale che facilita la scoperta di pattern o anomalie all'interno dei dati raccolti.

1.5.8 Data Visualization

La capacità di esaminare volumi significativi di informazioni e individuare intuizioni preziose potrebbe risultare inutile nel caso in cui soltanto gli analisti fossero in grado di interpretare i risultati ottenuti.

La fase di Data Visualization ha lo scopo di impiegare metodi e strumenti di rappresentazione grafica per trasmettere in maniera efficace le conclusioni dell'analisi, facilitando, quindi, un'interpretazione accurata da parte degli utenti aziendali.

Tali utenti devono possedere le competenze necessarie per comprendere i risultati, in modo da trarre beneficio dall'analisi e, successivamente, essere in grado di fornire un feedback utile.

Al termine di questo stadio, gli utenti acquisiscono la capacità di svolgere analisi visive, consentendo loro di scoprire risposte a domande che potrebbero non aver ancora preso in considerazione.

I risultati in questione possono essere espressi attraverso diverse modalità, influenzando, di conseguenza, la loro interpretazione. Pertanto, risulta cruciale adottare le tecniche di visualizzazione più idonee, tenendo conto del contesto aziendale specifico.

Un ulteriore aspetto da valutare consiste nell'offrire un metodo di approfondimento (drill-down) per confrontare statistiche elementari, in modo da permettere agli utenti di comprendere il processo che ha generato i risultati aggregati.

1.5.9 Utilization of Analysis Results

In questa fase si esplora la determinazione delle modalità e delle aree in cui i risultati ottenuti dall'analisi dei dati possono essere ulteriormente impiegati. A seconda delle specificità del problema affrontato, l'analisi potrebbe generare modelli che offrono nuove conoscenze e permettono di individuare schemi o relazioni tra vari fattori. Generalmente, un modello si presenta sotto forma di un'equazione matematica o di un insieme di regole.

Questi modelli possono essere utilizzati per perfezionare la logica dei processi aziendali e dei sistemi applicativi, contribuendo, così, all'ottimizzazione delle prestazioni complessive. Ne consegue che i risultati analitici si possono integrare, sia automaticamente che manualmente, nei sistemi aziendali per migliorarne l'efficienza e l'efficacia.

Inoltre, l'identificazione di pattern, correlazioni e anomalie può essere sfruttata per affinare e perfezionare i processi aziendali, agendo in maniera mirata sulla base delle informazioni ricavate dall'analisi. In tal modo, i risultati analitici diventano un valido strumento per l'ottimizzazione e il perfezionamento dei processi di business.

CAPITOLO 2

Introduzione a Power BI

Nel presente capitolo ci proponiamo di esaminare e approfondire il software di Business Intelligence, Power BI, strumento ampiamente utilizzato per l'analisi e la visualizzazione dei dati. Analizzeremo l'architettura di Power BI, con particolare attenzione al componente chiave del sistema, Power BI Desktop. Inoltre, discuteremo il processo di data cleaning all'interno di Power BI, esplorando le funzionalità di Power Query e la sua importanza nell'elaborazione dei dati. Infine, ci concentreremo sulla componente di Data Visualization del software, indagando i diversi tipi di visualizzazione offerti, l'uso di filtri e l'implementazione delle misure DAX per migliorare e personalizzare l'analisi dei dati.

2.1 Che cos'è Power BI?

Power BI è uno strumento di Business Intelligence sviluppato da Microsoft. Il suo scopo principale è supportare l'analisi dei dati aziendali attraverso rappresentazioni visive interattive e un'interfaccia user-friendly. Ciò consente agli utenti di elaborare in modo accurato e rapido report e dashboard informativi.

Power BI fornisce una suite di servizi software, applicazioni e connettori progettati per trasformare dati provenienti da fonti eterogenee in informazioni coerenti, graficamente attraenti e intuitive. Questi dati possono includere fogli di calcolo Excel, raccolte di data warehouse basate sul cloud, o configurazioni ibride locali, tra le altre opzioni. Power BI facilita la connessione a queste fonti di dati, la visualizzazione e l'individuazione delle informazioni rilevanti e la condivisione di tali informazioni con gli utenti interessati.

Power BI è un prodotto software relativamente recente che ha rapidamente guadagnato terreno nel settore della Business Intelligence, grazie agli ingenti investimenti da parte di Microsoft e alla sua capacità di competere con gli attuali leader di mercato, come QlikSense e Tableau. Il costante rilascio di aggiornamenti mensili e l'enfasi sul miglioramento continuo hanno contribuito al posizionamento di Power BI come leader nel Magic Quadrant di Gartner per gli strumenti di analisi dei dati.

2.2 Architettura

Power BI è composto da vari elementi che interagiscono tra loro (Figura 2.1); gli elementi chiave sono:

- *Power BI Desktop*: un'applicazione desktop per Windows dedicata alla creazione, modifica e visualizzazione di report, che mette a disposizione degli utenti strumenti per analizzare e gestire i dati.
- *Power BI Service*: un servizio online basato sul modello Software as a Service (SaaS) che consente di visualizzare e condividere dashboard con informazioni aggiornate in tempo reale, facilitando la collaborazione e il monitoraggio delle performance aziendali.
- *Gateway di Power BI*: i gateway servono per sincronizzare il flusso dei dati in ingresso e in uscita all'interno del sistema Power BI nella maniera più efficiente ed efficace possibile.
- *Power BI Mobile*: comprende le applicazioni per dispositivi Android e iOS, offre la possibilità di monitorare e accedere ai dati creati tramite Power BI Desktop da qualsiasi dispositivo mobile, garantendo flessibilità e accesso alle informazioni aziendali in ogni momento.

Oltre a questi quattro componenti, Power BI ne comprende altri due:

- *Power BI Report Builder*, destinato alla realizzazione di report impaginati che verranno successivamente condivisi tramite il servizio Power BI.
- *Power BI Report Server*, un server di reportistica locale che consente la pubblicazione dei report elaborati tramite Power BI Desktop.

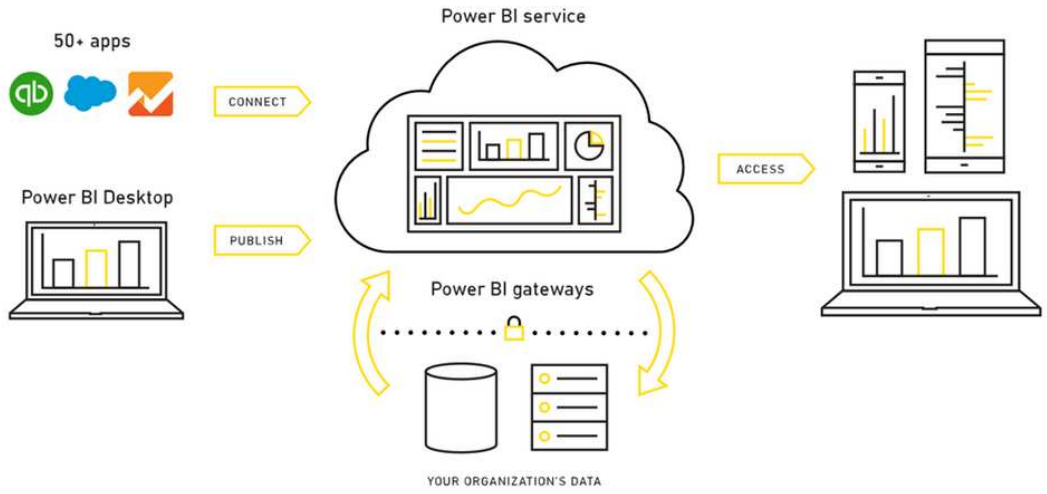

Figura 2.1: Architettura di Power BI

2.3 Power BI Desktop

Il nostro studio si focalizzerà sull'utilizzo di Power BI Desktop, uno strumento potente che ci permette di importare i dati da moltissime tipologie di sorgenti, al fine di elaborare ed estrarre informazioni utili.

Il flusso di azioni in Power BI solitamente comprende:

- Stabilire connessioni con diverse fonti di dati.

- Effettuare operazioni di data shaping mediante query al fine di creare modelli di dati efficaci.
- Impiegare tali modelli per generare visualizzazioni e report.
- Condividere i file dei report con altri utenti, consentendo loro di utilizzarli, espanderli e distribuirli. Sebbene sia possibile condividere i file con estensione `pbix` come qualsiasi altro file, la modalità più efficiente consiste nel caricarli nel servizio Power BI.

Power BI Desktop integra la solida tecnologia di Microsoft Query Engine con le funzionalità di modellazione dei dati e visualizzazione. Di conseguenza, gli analisti dei dati e altri utenti possono sviluppare e distribuire facilmente collezioni di query, modelli e report. La sinergia tra Power BI Desktop e il servizio Power BI consente di gestire, elaborare, condividere e ampliare informazioni dettagliate derivate dai dati in modo più agevole.

In sintesi, Power BI Desktop facilita e ottimizza il processo di progettazione e creazione di repository e report di Business Intelligence, che altrimenti risulterebbe disordinato, frammentato e complesso.

La schermata iniziale di Power BI Desktop fornisce tre tipologie di viste:

- *Report*: permette all'utente di sfruttare le query create al fine di produrre rappresentazioni grafiche e organizzarle su una o più pagine.
- *Dati*: offre la possibilità di esaminare le informazioni caricate nel report attraverso un formato di modello di dati, al quale è possibile aggiungere misure, creare colonne aggiuntive e controllare le relazioni esistenti.
- *Relazioni*: consente di visualizzare, gestire e modificare, se necessario, una rappresentazione grafica delle relazioni definite all'interno del modello di dati.

Le tre modalità di visualizzazione in Power BI possono essere attivate facendo clic sulle corrispondenti icone posizionate lungo il bordo sinistro dell'interfaccia. Nella Figura 2.2, si può notare come sia stata selezionata la *Visualizzazione Report*.

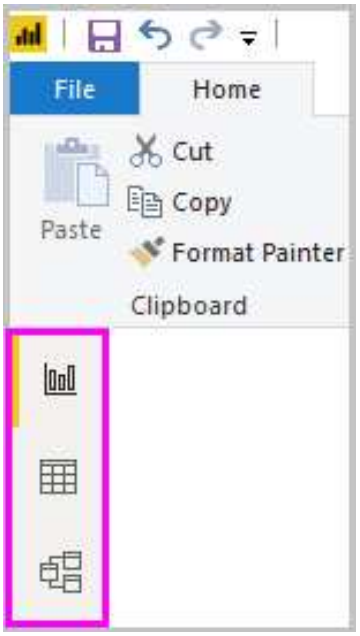

Figura 2.2: Modalità di visualizzazione in Power BI Desktop

2.4 Data Cleaning

Una volta completata l'installazione di Power BI Desktop, è possibile accedere all'ampia gamma di dati in costante crescita. Per esplorare le varie fonti di dati disponibili, è necessario selezionare "*Recupera dati*" > "*Altro*" nella sezione *Home* di Power BI Desktop e, successivamente, nella finestra "*Recupera dati*", sfogliare l'elenco completo delle fonti di dati disponibili (Figura 2.3).

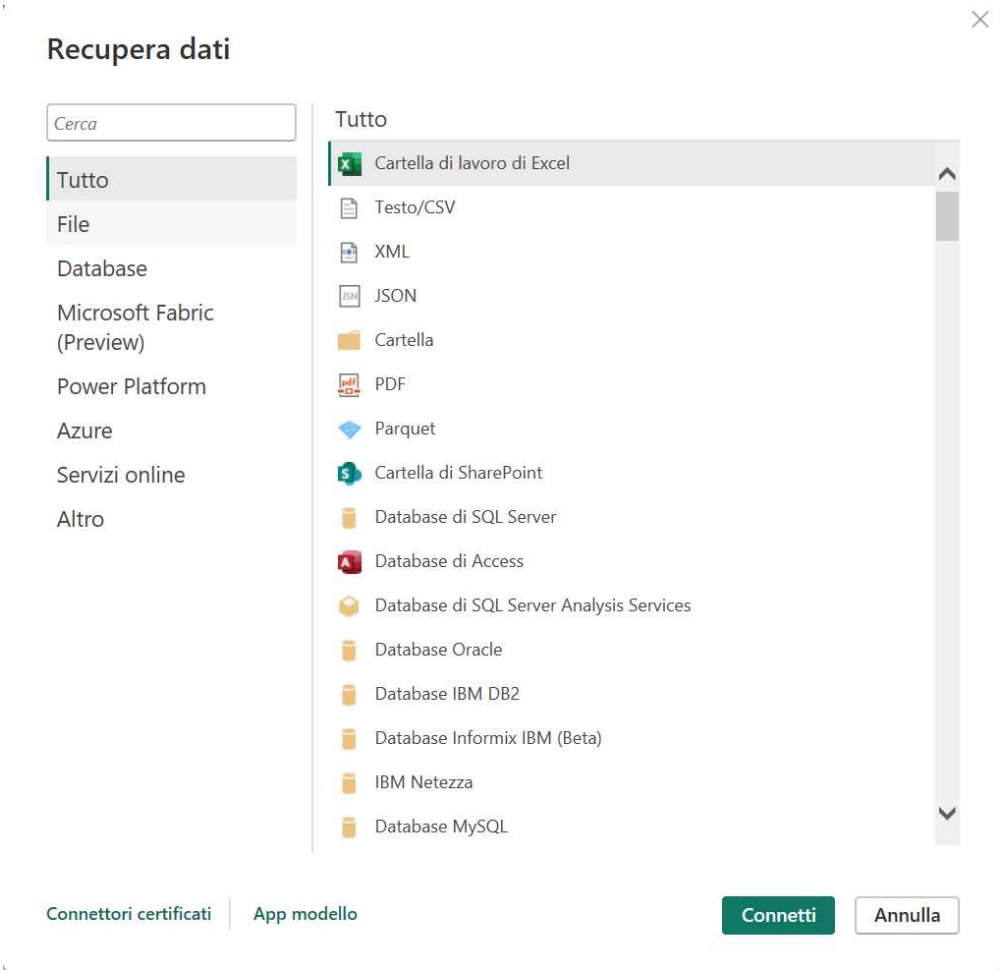

Figura 2.3: Finestra *Recupera dati* in Power BI Desktop

Stabilito il collegamento con la fonte di dati, si procede con la fase di trasformazione e pulizia degli stessi. Questa fase riveste un'importanza cruciale e influisce sulla qualità delle analisi basate sui dati stessi. Alcuni interventi tipici durante questo step includono l'eliminazione delle ripetizioni, la sostituzione dei valori, la rimozione di dati inesatti e l'adeguamento dei tipi di dati.

2.4.1 Power Query

Il software Power BI Desktop integra anche il componente Power Query Editor, il quale si apre in una finestra distinta. Tramite l'Editor di Power Query, è possibile generare interrogazioni e manipolare i dati, per poi importare il modello di dati perfezionato all'interno di Power BI Desktop al fine di elaborare report.

Dopo aver stabilito la connessione, apparirà la finestra dello *Strumento di navigazione* come mostrato nella Figura 2.4.

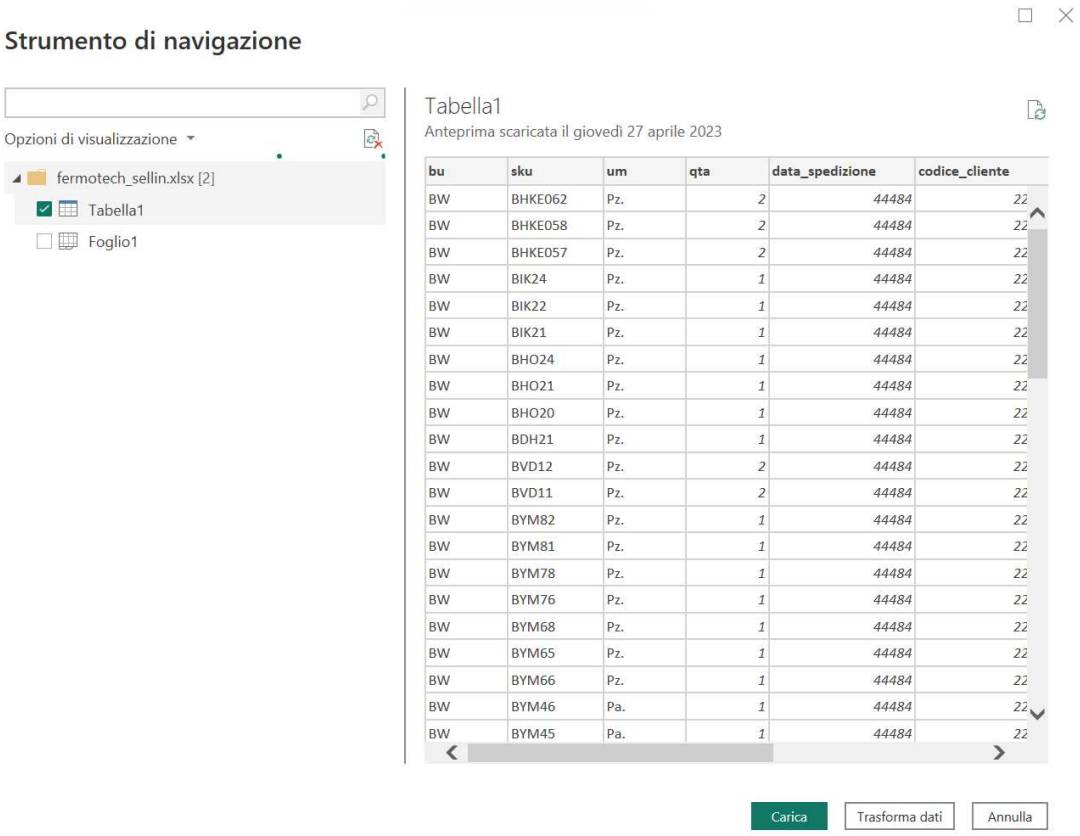

Figura 2.4: Finestra *strumento di navigazione* in Power BI Desktop

In questa finestra, è possibile visualizzare un'anteprima dei dati. Si ha la possibilità di scegliere "*Carica*" per importare direttamente la tabella, oppure "*Trasforma dati*" per apportare modifiche alla tabella prima di caricarla. Selezionando "*Trasforma dati*", l'editor di Power Query si avvierà mostrando una vista rappresentativa della tabella. Nel lato destro della finestra possiamo trovare il pannello "*Impostazioni query*", come mostrato nella Figura 2.5, accessibile anche dalla scheda "*Visualizza*" all'interno dell'editor di Power Query.

A questo punto è possibile modificare i dati in base alle proprie esigenze, o meglio sottoporli a data shaping. È importante fornire precise direttive all'editor di Power Query per la modifica dei dati nel corso dell'importazione e della visualizzazione. Ciò non comporterà alcuna alterazione della fonte dei dati originale, ma unicamente la modifica di questa vista specifica.

Il data shaping implica trasformazioni dei dati, come rinominare colonne o tabelle, eliminare righe o colonne, o modificare i tipi di dati. L'Editor di Power Query registra tali operazioni in sequenza nell'elenco "*Passaggi applicati*" presente nel riquadro delle Impostazioni Query. Quando la query si connette all'origine dei dati, questi passaggi vengono eseguiti sistematicamente per garantire che i dati siano sempre conformi alla struttura specificata. Tale processo si verifica durante l'utilizzo della query in Power BI Desktop, così come quando altri utenti accedono alla query condivisa, grazie all'utilizzo di Power BI Service.

2.5 Data Visualization

La Data Visualization consiste nell'utilizzo di elementi visivi, quali grafici, diagrammi, infografiche, e, talvolta, anche animazioni, al fine di mostrare informazioni estrapolate

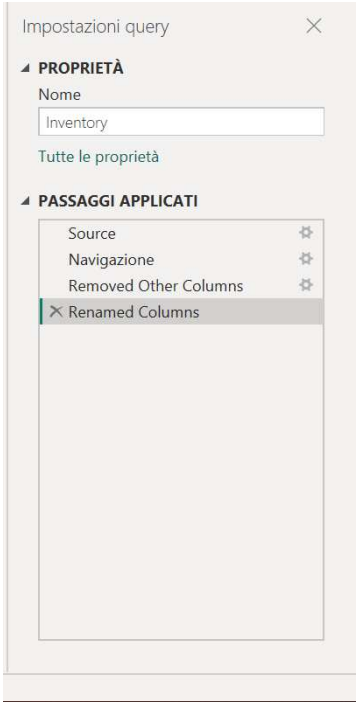

Figura 2.5: Pannello *Impostazioni query* in Power BI Desktop

dai dati. Questo metodo consente di comunicare in maniera semplice e chiara le relazioni complesse tra i dati e le conoscenze ad essi associate.

Power BI consente una visualizzazione dei dati rapida ed efficace, fornendo un'ampia gamma di strumenti per migliorare questa fase. Tra questi strumenti ritroviamo:

- numerosi tipi di visualizzazione;
- filtri;
- Data Analysis eXpressions.

Nei paragrafi successivi, esamineremo in profondità ciascuno di questi elementi.

2.5.1 Tipi di visualizzazione

Una volta realizzato un modello dei dati, si può procedere ad inserire i vari campi all'interno dell'area di disegno della vista Report, dando vita a specifiche rappresentazioni grafiche delle informazioni presenti nel modello stesso. Tali rappresentazioni sono note come oggetti visivi. In determinate situazioni, potrebbe essere necessario produrre un insieme di oggetti visivi aventi lo scopo di fornire un'analisi completa dei dati impiegati nella creazione del modello tramite Power BI. Tale insieme di oggetti visivi prende il nome di report. La Figura 2.6 mostra un esempio delle diverse visualizzazioni disponibili all'interno di Power BI. Alcuni esempi di visualizzazione utilizzati di frequente sono:

- *Grafici ad area di base*: il grafico ad area di base si fonda sul concetto di grafico a linee, con l'aggiunta del riempimento dell'area tra l'asse e la linea stessa. Questo tipo di grafico è particolarmente utile per mettere in evidenza le variazioni nel tempo e per focalizzare l'attenzione sul valore complessivo di una tendenza. Un esempio di applicazione può essere la rappresentazione del profitto nel tempo, utilizzando un grafico ad area per sottolineare l'intero ammontare del guadagno.

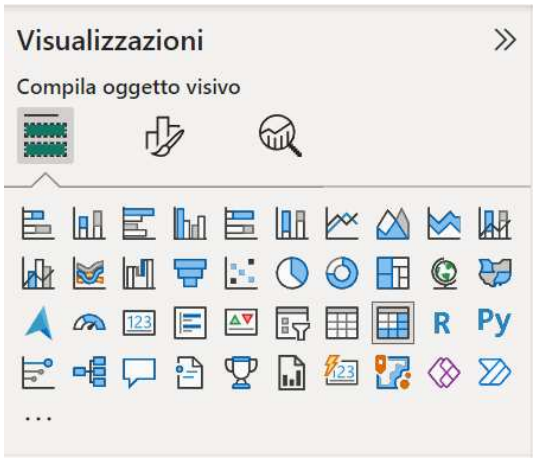

Figura 2.6: Pannello *visualizzazioni* in Power BI Desktop

- *Istogrammi*: questo tipo di grafico illustra una distribuzione di valori numerici mediante l'uso di un diagramma a barre (senza alcuna distanza tra le barre), che rappresenta la quantità di dati inclusi in un specifico intervallo. Tale rappresentazione grafica permette all'osservatore di riconoscere agevolmente eventuali anomalie presenti all'interno di un insieme di dati analizzato.
- *Alberi di scomposizione*: l'albero di scomposizione offre la possibilità di esaminare le informazioni da diverse prospettive. Attraverso l'aggregazione automatizzata dei dati, permette di eseguire il drill-down nelle dimensioni in qualsiasi ordine desiderato. Inoltre, si tratta di un dispositivo di visualizzazione che impiega l'Intelligenza Artificiale, al quale può essere richiesto di identificare la dimensione successiva da analizzare in base a determinati parametri. Di conseguenza, l'albero di scomposizione si rivela un prezioso strumento per esplorazioni ad hoc e ricerche di analisi delle cause principali.
- *Grafici a torta*: i grafici a torta indicano la relazione tra un intero e le parti.
- *Schede con numero singolo*: le schede con numero singolo mostrano un dato isolato, un'unica informazione numerica. In alcune occasioni, la finalità principale di un report di Power BI è esporre un unico valore, come, ad esempio, le vendite complessive.
- *Schede con più righe*: le schede a righe multiple espongono diversi elementi informativi, rappresentando ciascun valore numerico su righe distinte, permettendo, così, una visualizzazione chiara e ordinata di più punti dati.

2.5.2 Filtri

Una delle caratteristiche fondamentali di Power BI Desktop è la capacità di filtrare i dati. Durante la creazione di un report, è possibile applicare filtri diversi per migliorare la rappresentazione grafica delle informazioni. Nel pannello "*Filtri*" possiamo distinguere tre tipi di filtri che ci permettono di esaminare accuratamente le visualizzazioni, ovvero:

- *Filtro per pagina*: influisce su tutti gli elementi visivi presenti nella pagina del report.
- *Filtro per elemento visivo*: riguarda un singolo elemento visivo su una pagina del report.
- *Filtro per l'intero report*: si estende a tutte le pagine del report.

Il pannello "*Filtri*" consente anche di utilizzare un campo che ancora non è incluso nei vari elementi grafici del report. Ogni filtro può avere sotto-filtri, utili per affinare ulteriormente l'analisi dei dati. Tra i principali tipi di sotto-filtri troviamo:

- *Filtro avanzato*: permette di filtrare i dati applicando condizioni specifiche (ad esempio, dimensione, maggiore o minore di, uguale a, etc.).
- *Filtro basilare*: adatto quando si desidera selezionare valori particolari (ad esempio, anno, mese, giorno).
- *Filtro Top N*: utile quando si devono eseguire analisi su determinati valori.

Attraverso queste numerose opzioni di filtraggio, Power BI Desktop facilita la comprensione e l'analisi dei dati, permettendo di personalizzare e approfondire la visualizzazione delle informazioni.

2.5.3 Data Analysis eXpressions

Durante l'analisi dei dati in Power BI vengono eseguiti calcoli personalizzati, le cosiddette *misure*, necessarie per esaminare i dati in modo aggregato e fornire informazioni utili. Le misure possono essere utilizzate nelle visualizzazioni di Power BI come elementi di dati nei grafici, nelle tabelle o in qualsiasi altra visualizzazione che supporti il consumo di misure. Le misure si basano sull'elaborazione di risultati attraverso specifiche espressioni matematiche. Durante la creazione di misure personalizzate, si utilizza il linguaggio DAX (Data Analysis eXpressions), noto per la sua vasta libreria di oltre 200 funzioni, operatori e strutture. Grazie a tale libreria, è possibile ottenere una grande flessibilità nella generazione di misure e nell'elaborazione di espressioni di calcolo.

Le formule DAX mostrano somiglianze con le formule di Excel e condividono molte funzioni comuni, come DATE, SUM e LEFT. Tuttavia, le funzioni DAX sono progettate per essere impiegate in un contesto di dati relazionale, tipici di Power BI Desktop. Pertanto, l'utilizzo di DAX è particolarmente indicato per lavorare con le informazioni in questa piattaforma.

Durante l'elaborazione di formule tramite Data Analysis Expressions, è fondamentale prestare attenzione al tipo di dato per prevenire incoerenze e incongruenze nei risultati ottenuti. I tipi di dato compatibili includono numeri interi, decimali, valute, booleani, testo e date. Inoltre, nelle espressioni DAX, è necessario specificare le tabelle e le colonne in modo appropriato (utilizzando apici nel caso in cui il nome della tabella contenga spazi). Per racchiudere colonne e misure, è opportuno impiegare le parentesi quadre. È importante ricordare alcuni aspetti cruciali durante l'elaborazione di formule ed espressioni DAX:

- le formule e le espressioni DAX non hanno la capacità di alterare o inserire valori individuali all'interno delle tabelle;
- con DAX, non è possibile generare righe calcolate; tuttavia, è consentita la creazione di colonne e misure;
- durante la definizione delle colonne in DAX, è possibile annidare funzioni a qualsiasi livello, garantendo una maggiore flessibilità nella gestione delle espressioni.

CAPITOLO 3

Descrizione dei dati a disposizione e attività di ETL

In questo capitolo esploreremo i vari tipi di dati: strutturati, non strutturati, semistrutturati e metadati, evidenziando il loro ruolo chiave nell'analisi delle informazioni. Focalizzeremo l'attenzione sui dataset forniti dall'azienda Bros Manifatture, per poi approfondire il processo di Estrazione, Trasformazione e Caricamento (ETL), fondamentale nel campo del data warehousing. Infine, presenteremo un dettagliato resoconto delle operazioni di ETL condotte sui nostri dataset attraverso l'uso del software Power BI.

3.1 Introduzione sui tipi di dati

Elementi fondamentali per la realizzazione di una campagna di Data Analytics sono i dati. Un report dell'IBM Marketing Cloud, intitolato "10 Trend Fondamentali di Marketing per il 2017", ha evidenziato un dato allarmante: il 90% dell'intera mole di dati attualmente presenti nel mondo è stato prodotto solamente negli ultimi anni, con una media impressionante di 2.5 quintilioni di byte al giorno. Stiamo assistendo a un'esplosione di dati senza precedenti; l'uso incessante di internet e dei dispositivi tecnologici sta digitalizzando e trasformando in dati ogni aspetto delle nostre vite quotidiane. Questa la previsione per il futuro: il volume dell'universo digitale è destinato a raddoppiare almeno ogni due anni, con una espansione 50 volte superiore rispetto al decennio 2010-2020. I dati generati sia da esseri umani sia da dispositivi tecnologici crescono a un ritmo dieci volte superiore rispetto ai dati tradizionali di carattere aziendale, e i dati generati dalle macchine aumentano a un ritmo addirittura 50 volte superiore (Figura 3.1).

Le informazioni prodotte dall'uomo e quelle generate dalle macchine provengono da un'ampia gamma di fonti e possono essere illustrate in diversi formati o categorie. Di seguito, analizzeremo i vari tipi di dati che le soluzioni di Big Data sono in grado di processare. Le principali categorie di dati includono:

- dati strutturati;
- dati non strutturati;
- dati semi-strutturati.

Questi tre tipi di dati riguardano la struttura interna e in alcuni casi sono denominati formati dei dati. Oltre a questi tipi di dati, nel campo dei Big Data, sono rilevanti anche i *metadati*.

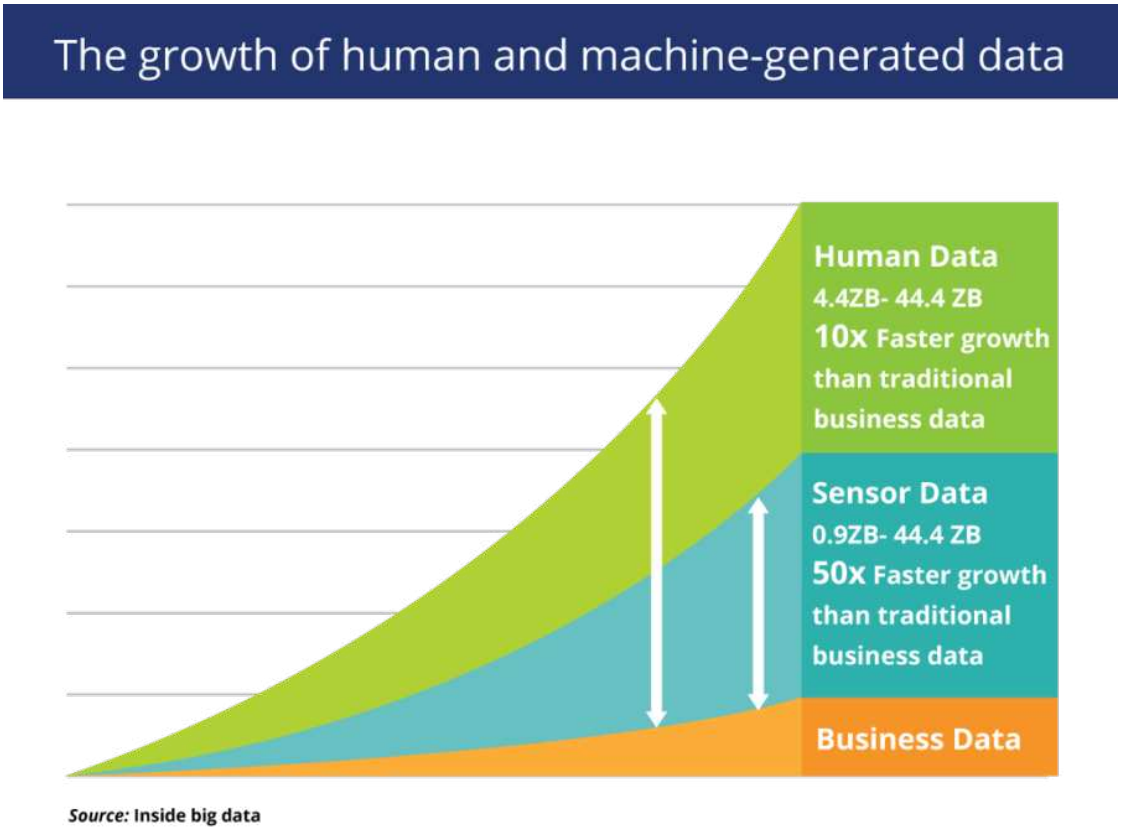

Figura 3.1: Crescita dei dati generati dagli umani e dalle macchine

3.1.1 Dati strutturati

I dati strutturati si riferiscono a informazioni che aderiscono a un particolare modello o schema. Queste informazioni sono tipicamente organizzate su tabelle per evidenziare le relazioni tra diverse entità, motivo per cui vengono frequentemente depositate in un database relazionale. Questo formato di dati è comunemente prodotto da applicazioni aziendali e sistemi informativi, come i sistemi di pianificazione delle risorse aziendali (ERP) e i sistemi di gestione delle relazioni con i clienti (CRM). Grazie all'elevata disponibilità di tool e database che supportano intrinsecamente i dati strutturati, questi non necessitano di considerazioni speciali per quanto riguarda l'elaborazione o la conservazione.

3.1.2 Dati non strutturati

I dati non strutturati sono informazioni che non rispettano un preciso schema o modello di dati. Si calcola che in un'organizzazione tipica, fino all'80% dei dati appartengono a questa categoria. La quantità di dati non strutturati tende a crescere a un ritmo più elevato rispetto ai dati strutturati.

Queste informazioni possono essere di tipo testuale o binario e sono spesso memorizzate in file indipendenti che non presentano relazioni intrinseche tra di loro. Un file di testo ha la capacità di conservare materiali da varie fonti, come tweet e articoli di blog. I file binari, invece, sono normalmente file multimediali che contengono informazioni sotto forma di immagini, suoni o video. Nonostante sia i file di testo che quelli binari abbiano una struttura determinata dal formato del file stesso, la caratteristica distintiva dei dati non strutturati non riguarda la struttura del file, bensì il formato dei dati conservati al loro interno.

Per gestire e conservare dati non strutturati, è spesso richiesta una logica soggetta al contesto. Ad esempio, per riprodurre un file video, è fondamentale disporre del *codec (codificatore-decodificatore)* appropriato. I dati non strutturati non possono essere direttamente processati o richiesti tramite SQL. Se si devono conservare tali dati all'interno di un database relazionale, essi vengono memorizzati in una tabella strutturata come BLOB (Binary Large Object). In alternativa, esistono anche i database Not-only-SQL (NoSQL), che rappresentano una tipologia di database non relazionale. Questi ultimi sono particolarmente idonei per il deposito di dati non strutturati, in quanto possono memorizzare dati non strutturati insieme a dati di natura strutturata. Nella Figura 3.2 vengono evidenziate le principali differenze tra dati strutturati e dati non strutturati.

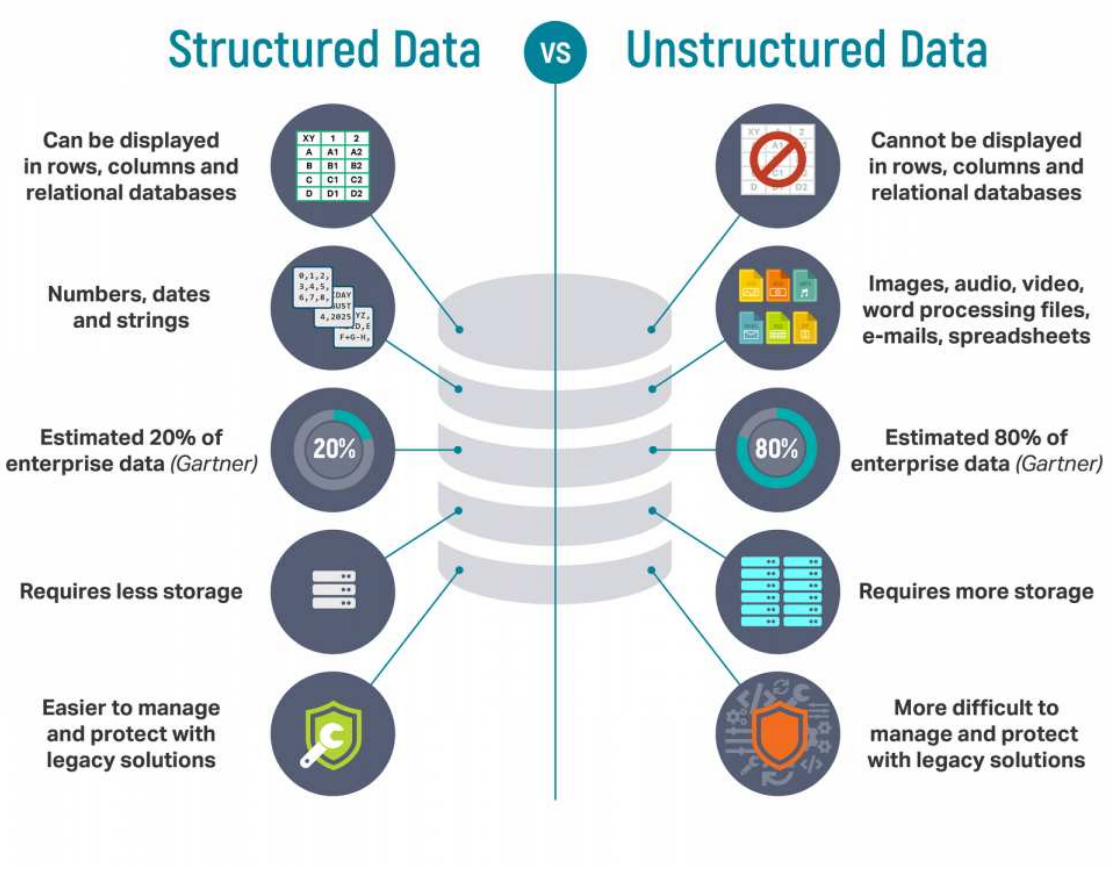

Figura 3.2: Differenze fondamentali tra dati strutturati e dati non strutturati

3.1.3 Dati semi-strutturati

I dati semi-strutturati presentano una determinata forma strutturale e consistenza; tuttavia, non sono per definizione relazionali. Al contrario, questi dati mantengono una natura gerarchica o basata sul modello a grafo. Tali informazioni vengono prevalentemente conservate in documenti che contengono testo. Ad esempio, i file XML e JSON sono alcune delle rappresentazioni più comuni di dati semi-strutturati. Data la loro natura testuale e l'adesione a una specifica struttura, la loro elaborazione risulta più semplice rispetto a quella dei dati totalmente non strutturati.

Tipiche fonti di dati semi-strutturati comprendono file EDI (Electronic Data Interchange), fogli di calcolo, feed RSS e informazioni provenienti da sensori. Questo tipo di dati richiede, spesso, considerazioni particolari per quanto riguarda la loro elaborazione preliminare e

la loro memorizzazione, specialmente quando il formato di base non è testuale. Un'istanza specifica di pre-elaborazione di dati semi-strutturati è la validazione di un file XML. Questo processo è necessario per garantire che il file sia in linea con la definizione del suo schema.

3.1.4 Metadati

I metadati giocano un ruolo fondamentale nell'ambiente moderno dei dati; essi, infatti, forniscono una descrizione delle caratteristiche e dell'architettura di un dataset. Gran parte di queste informazioni, che sono create da dispositivi e software automatici, vengono abitualmente associate ai dati a cui si riferiscono.

È di fondamentale importanza tracciare e gestire accuratamente i metadati durante l'elaborazione, l'archiviazione e l'analisi dei Big Data. Questo perchè i metadati offrono elementi preziosi riguardo all'origine e alla genealogia dei dati, rendendoli particolarmente utili per comprendere il contesto e la validità delle informazioni durante l'elaborazione.

I metadati possono assumere varie forme e possono riguardare aspetti diversi delle informazioni. Ad esempio, possono contenere tag XML che indicano l'autore di un documento e la data in cui è stato generato. Essi possono, anche, includere attributi che riportano dettagli utili, come la dimensione di un file o la risoluzione di un'immagine digitale.

Le soluzioni di Big Data dipendono notevolmente dai metadati, soprattutto quando devono processare informazioni semi-strutturate o non strutturate. Questo perché i metadati possono aiutare a imporre una qualche forma di struttura su questi dati, facilitandone l'elaborazione e l'analisi. Infine, va notato che il corretto utilizzo dei metadati può contribuire ad ottimizzare l'accessibilità dei dati, rendere più efficace il recupero delle informazioni e migliorare la gestione del ciclo di vita dei dati.

3.2 I Dataset sulle vendite di Bros

In questo capitolo introduciamo una serie di dataset che ci sono stati forniti da Bros Manifatture, azienda che opera da oltre quarant'anni nel settore della gioielleria ed orologeria. Esamineremo dettagliatamente i dataset relativi alle vendite di Bros alle attività commerciali e quelli relativi alle vendite al consumatore finale. Le informazioni a nostra disposizione si riferiscono ai prodotti venduti nell'ultimo trimestre 2021, nell'anno 2022 e nel primo trimestre 2023. La fonte predominante di queste informazioni è Bros Manifatture; essendo questi dati interni, presentano una minore interferenza rispetto alle informazioni provenienti da diverse fonti. L'azienda ha messo a disposizione questi dataset condividendoli su un repository di Google Drive. Disponiamo della possibilità di accedere ai dati nel formato Excel, adatto per condurre le nostre analisi, come evidenziato nella Figura 3.3.

Il fulcro del nostro caso di studio è l'analisi dei dati di vendita di Bros Manifatture dall'ultimo trimestre 2021 al primo trimestre 2023. Questa analisi ci consentirà di delineare e studiare i trend di vendita, i brand più venduti, i clienti più operativi, e molte altre peculiarità. Successivamente, questi dati analizzati ci agevoleranno nello sviluppo di report interattivi di facile comprensione e nell'estrazione di informazioni rilevanti per guidare l'azienda nelle decisioni strategiche.

Nelle prossime sezioni ci soffermeremo sull'esame dettagliato dei dataset a disposizione. Questi dataset sono relazionali; pertanto i dati in essi presenti sono strutturati. Entrambi i dataset sono formati da una tabella che rappresenta la base di dati principale. La base dati principale consiste in una lista di articoli venduti; quindi, ogni riga rappresenta un unico prodotto con distinti attributi, i quali verranno esaminati dettagliatamente nelle sezioni successive.

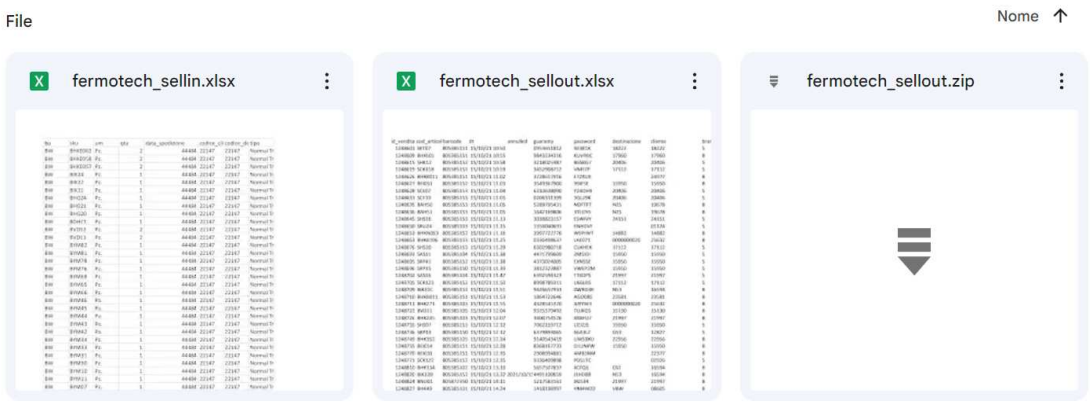

Figura 3.3: Schermata di Google Drive contenente i vari dataset a disposizione

3.2.1 Bros: dataset Sell Out

Il dataset "*Sell Out*" contiene le informazioni relative alle vendite al consumatore finale nel periodo compreso tra l'ultimo trimestre 2021 e il primo trimestre 2023. La tabella è strutturata in 11 colonne, ciascuna delle quali rappresenta un tipo diverso di informazione. Essa comprende più di 350000 righe, ciascuna riga contiene i dati relativi a una specifica unità di prodotto venduta al consumatore finale. In sostanza, ogni articolo venduto è caratterizzato da una serie di 11 attributi distintivi, i più rilevanti dei quali sono i seguenti:

- *Id_vendita*: questo attributo indica il codice univoco associato alla singola vendita di un articolo. Un esempio è 2102998.
- *Cod_articolo*: questo attributo racchiude il codice univoco utilizzato da Bros Manifatture per identificare ogni prodotto. Un esempio è BEI019.
- *Barcode*: anche questo attributo, come il campo *Cod_articolo*, identifica in maniera univoca l'articolo venduto. Un esempio è 8053851501082.
- *Dt*: questo attributo si riferisce alla data e all'orario in cui un determinato prodotto è stato venduto al consumatore finale. Un esempio à 30/10/2021 18:58:48.
- *Annulled*: questo attributo detiene una dualità di valori, la data e l'ora del giorno di restituzione o sostituzione dell'articolo. Tuttavia, in assenza di tali eventi, l'attributo si presenta con un valore vuoto, conosciuto come valore nullo.
- *Cliente*: questo attributo indica il codice univoco che Bros Manifatture utilizza per identificare ogni cliente. Un esempio è 18222.
- *Brand*: questo attributo indica il brand a cui appartiene un determinato articolo. Un esempio è S.
- *Prezzo L20*: questo attributo si riferisce al prezzo dell'articolo. Un esempio è 24.

La Figura 3.4 mostra un estratto selezionato dal dataset "*SellOut*".

3.2.2 Bros: dataset Sell In

Il dataset "*SellIn*" contiene le informazioni relative alle vendite ai clienti nel periodo compreso tra l'ultimo trimestre 2021 e il primo trimestre 2023. La tabella è strutturata in 8

id_vendita	cod_articolo	barcode	dt	annulled	guaranty	password	destinazione	cliente	brand	PREZZO L20
1248601	SKT07	805385151547	15/10/21 10.50		0953651812	9Z381X	18222	18222	S	24
1248609	BHKL01	805385151700	15/10/21 10.55		9843234316	KUVRBC	17960	17960	B	34
1248615	SHK12	805385152343	15/10/21 10.58		3218025987	B6SBG7	20406	20406	S	18
1248619	SCK118	805385152151	15/10/21 10.59		3452908712	VNFI7P	17112	17112	S	19
1248626	BHKB011	805385151730	15/10/21 11.02		3728617916	E7Z4UX		24977	B	34
1248627	BEI051	805385152610	15/10/21 11.03		3549367900	99IF5E	15950	15950	B	48
1248628	SCE07	805385153236	15/10/21 11.04		6310638090	Y24OH9	20406	20406	S	24
1248633	SCE10	805385153241	15/10/21 11.05		0206331399	3GL29K	20406	20406	S	49
1248635	BAH50	805385153073	15/10/21 11.05		5289795431	NOFTFT	N25	19678	B	46
1248636	BAH51	805385153074	15/10/21 11.05		1642169806	3TLGY6	N25	19678	B	28
1248645	SHS16	805385153111	15/10/21 11.13		3038823157	ESWIVY	24151	24151	S	21
1248650	SRU24	805385103200	15/10/21 11.15		1356040691	ENHDVF		01124	S	24
1248653	BHKN053	805385152595	15/10/21 11.16		3997722776	W9PIWT	14882	14882	B	34
1248663	BHKB106	805385153382	15/10/21 11.25		0336498637	LAE071	0000000020	25632	B	44

Figura 3.4: Schermata contenente alcune righe del dataset "*SellOut*"

colonne, ognuna delle quali rappresenta un tipo diverso di informazione. Essa comprende più di 500000 righe; ciascuna riga contiene i dati relativi a un determinato prodotto spedito in diverse quantità al cliente. In sostanza, ogni tipologia di articolo spedito è caratterizzata da una serie di 8 attributi distintivi, i più rilevanti dei quali sono i seguenti:

- *Sku*: questo attributo racchiude il codice univoco utilizzato da Bros Manifatture per identificare ogni prodotto. Un esempio è BEI019.
- *Data_spedizione*: questo attributo si riferisce alla data in cui un determinato ordine è stato spedito al cliente. Un esempio è 44484.
- *Cod_cliente*: questo attributo indica il codice univoco che Bros Manifatture utilizza per identificare ogni cliente. Un esempio è 18222.
- *Bu*: questo attributo indica il brand a cui appartiene un determinato articolo. Un esempio è BW.
- *Qta*: questo attributo si riferisce alla quantità di prodotti, di una determinata tipologia, che è stata spedita al cliente. Un esempio è 40.

La Figura 3.5 mostra un estratto selezionato dal dataset "*SellIn*".

bu	sku	um	qta	data_spedizione	codice_cliente	codice_destinazione	tipo
BW	BHKE062	Pz.	2	44484	22147	22147	Normal Trade Control Sell
BW	BHKE058	Pz.	2	44484	22147	22147	Normal Trade Control Sell
BW	BHKE057	Pz.	2	44484	22147	22147	Normal Trade Control Sell
BW	BIK24	Pz.	1	44484	22147	22147	Normal Trade Control Sell
BW	BIK22	Pz.	1	44484	22147	22147	Normal Trade Control Sell
BW	BIK21	Pz.	1	44484	22147	22147	Normal Trade Control Sell
BW	BHO24	Pz.	1	44484	22147	22147	Normal Trade Control Sell
BW	BHO21	Pz.	1	44484	22147	22147	Normal Trade Control Sell
BW	BHO20	Pz.	1	44484	22147	22147	Normal Trade Control Sell
BW	BDH21	Pz.	1	44484	22147	22147	Normal Trade Control Sell
BW	BVD12	Pz.	2	44484	22147	22147	Normal Trade Control Sell
BW	BVD11	Pz.	2	44484	22147	22147	Normal Trade Control Sell
BW	BYM82	Pz.	1	44484	22147	22147	Normal Trade Control Sell
BW	BYM81	Pz.	1	44484	22147	22147	Normal Trade Control Sell

Figura 3.5: Schermata contenente alcune righe del dataset "*SellIn*"

3.3 Attività di ETL

In questa sezione discuteremo ed esamineremo dettagliatamente le operazioni di ETL (Figura 3.6) eseguite sui dataset, relativi alle vendite, forniti da Bros Manifatture. L'attività di ETL (Extract, Transform, Load) costituisce uno dei passaggi preliminari essenziali nell'ambito

dell'analisi dei dati. Essa è costituita da un insieme di funzioni di estrazione, trasformazione e caricamento dei dati da una o più sorgenti a un sistema di destinazione. La sua finalità risiede nella conversione dei dati non elaborati in informazioni operative per le funzioni di Business Intelligence.

L'abbreviazione "ETL" si riferisce alle lettere iniziali delle tre fasi, descritte come segue:

- *Extract*: nella fase di estrazione, i dati vengono raccolti da sorgenti diverse, come database, file, fogli di calcolo, e vengono predisposti per la successiva fase di elaborazione.
- *Transform*: durante la fase di trasformazione, i dati estratti vengono convertiti in un formato compatibile e compatto con il sistema di destinazione.
- *Load*: la fase di caricamento implica l'importazione dei dati trasformati nel sistema di destinazione, di solito un data warehouse, un data mart o un software di Business Intelligence.

Le attività di ETL sono fondamentali nella Business Intelligence, nella gestione, migrazione e integrazione dei dati e nella preparazione di questi ultimi per l'analisi. Esse offrono alle aziende la capacità di integrare dati provenienti da vari sistemi, trasformandoli in informazioni utili per prendere decisioni informate.

Nelle sezioni successive, esamineremo con maggiore dettaglio queste tre fasi.

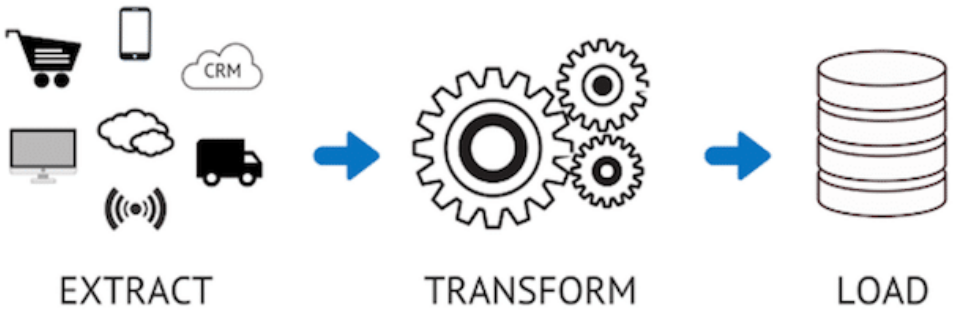

Figura 3.6: Attività di ETL

3.3.1 Extract

Nel processo di estrazione dei dati, il passo primario riguarda l'identificazione dei dataset che saranno soggetti ad analisi. Tipicamente, si tende a scegliere una pluralità di fonti di dati in modo da poter individuare schemi e collegamenti tra di essi. È doveroso sottolineare che le raccolte di dati possono essere interne all'organizzazione, ad esempio i database aziendali, oppure esterne, come i social media. Una volta stabilito quali set di dati sono pertinenti, si procede al recupero e alla memorizzazione degli stessi. Ciò assicura la disponibilità di una copia di tutti i dati indispensabili per l'analisi, i quali potranno essere modellati in modo ottimale successivamente.

L'estrazione può essere eseguita in vari modi tra cui sviluppare programmi su misura, utilizzare uno dei diversi strumenti di ETL disponibili sul mercato, o un mix di entrambi. Alcuni strumenti di ETL moderni possono semplificare enormemente le attività legate a questa fase, eliminando la necessità di scrivere righe di codice, poichè queste attività vengono gestite automaticamente da tali strumenti. Questo può ridurre notevolmente la quantità di personale necessario per questo step. Tuttavia, esiste un compromesso costi-benefici,

in quanto questi strumenti tendono ad essere più costosi vista la disponibilità gratuita di linguaggi di programmazione come Python.

3.3.2 Transform

In questa fase, vengono applicate varie funzioni o metodi per manipolare i dati estratti. Ciò è necessario per effettuare la pulizia dei dati, la combinazione dati da diverse fonti, la divisione in più tabelle o la combinazione di più tabelle in una, etc. L'obiettivo è di trasformare i dati in un formato che sia facilmente utilizzabile e che risponda meglio alle necessità aziendali.

Durante questa fase, il trattamento dei dati grezzi che vengono processati può includere le seguenti operazioni:

- rimozione di duplicati;
- validazione e pulizia dei dati per eliminare errori;
- mappatura dei dati nel formato richiesto;
- verifica dell'uniformità dei formati dei dati per assicurare la consistenza;
- filtraggio dei dati non pertinenti o superflui.

Inoltre, è possibile applicare step avanzati di trasformazione dei dati a seconda delle esigenze. Ad esempio:

- condensazione dei dati per ridurre le dimensioni del dataset;
- divisione dei dati in più colonne;
- unione di tabelle;
- codifica dei dati al fine di rispettare la normativa sulla riservatezza.

Attraverso questi step di trasformazione, quello che inizialmente rappresentava un insieme di materiale inutilizzabile viene trasformato in un prodotto di dati pronto per la fase finale del procedimento di ETL, ovvero quella del caricamento.

3.3.3 Load

Una volta completate le fasi di estrazione e trasformazione dei dati, l'ultima operazione consiste nel caricare i dati trasformati nel data warehouse. Questo processo è ben preciso, continuo e automatizzato; generalmente, il caricamento dei dati avviene in blocchi. Le modalità di caricamento dei dati sono le seguenti:

- *Caricamento totale dei dati*: questo tipo di caricamento si verifica solitamente durante la fase iniziale. Riguarda l'estrazione e la trasformazione del set di dati dalla fonte al data warehouse.
- *Caricamento incrementale dei dati*: questo processo consiste nel caricare periodicamente solo i dati aggiornati tra il sistema di origine e il sistema di destinazione.
- *Caricamento dei dati a blocchi*: se il set di dati è troppo grande, i dati vengono caricati in blocchi e periodicamente.
- *Caricamento incrementale in streaming*: questo metodo prevede lo streaming continuo dei dati, ma è adatto solo per set di dati più piccoli.

3.4 ETL sui dataset Bros

In questa sezione mostriamo dettagliatamente le procedure di Estrazione, Trasformazione e Caricamento applicate ai dataset di Bros Manifatture. In particolare, presentiamo un esame dettagliato delle tre fasi che includono numerosi screenshot per illustrare i vari passaggi, le interfacce con cui siamo entrati in contatto e le azioni intraprese. Abbiamo posto particolare attenzione sulle attività di pulizia e trasformazione dei dati. Queste attività sono state eseguite utilizzando l'editor di query di Power BI, denominato Power Query.

Gli insiemi di dati relativi alle vendite di Bros Manifatture hanno origine dalla medesima fonte ed hanno una struttura simile. Di conseguenza, le operazioni eseguite su entrambi sono in gran parte speculari. Per evitare ripetizioni inutili e per illustrare efficacemente la procedura adottata, abbiamo deciso di utilizzare il dataset "*SellOut*" come principale esempio di riferimento.

3.4.1 Extract

Come precedentemente menzionato, l'estrazione rappresenta la prima fase del processo di ETL. In questa attività facciamo uso di Power BI che ci fornisce un'ampia scelta di opzioni. L'interfaccia principale del software ci mostra ciò che è illustrato nella Figura 3.7, e cioè quattro delle opzioni più usate, insieme alla possibilità di altre selezioni. Questo si riflette anche nel contenuto della Figura 3.8, posizionato nella parte superiore del software Power BI. Utilizzando entrambe le opzioni, accediamo alla schermata di raccolta dei dati, come mostrato nella Figura 3.9. Qui abbiamo la possibilità di scegliere la nostra fonte di estrazione da una gamma di risorse compatibili con Power BI. Nel nostro caso di studio, useremo un foglio di lavoro di Excel, precedentemente convertito in tabella. Una volta effettuata la scelta, procediamo alla connessione tra la fonte selezionata e Power BI utilizzando la funzione "*Connetti*".

Figura 3.7: Schermata iniziale di Power BI per la selezione delle sorgenti dati

Dopo aver stabilito la connessione, si aprirà la schermata dello strumento di navigazione (Figura 3.10), dove possiamo visualizzare in anteprima i nostri dati, effettuare un'analisi preliminare e decidere se trasformarli o caricarli direttamente a condizione che siano già pronti per l'analisi. Nel nostro scenario, optiamo per la trasformazione prima del caricamento; dunque inizieremo con Power Query, come illustrato nella Figura 3.11.

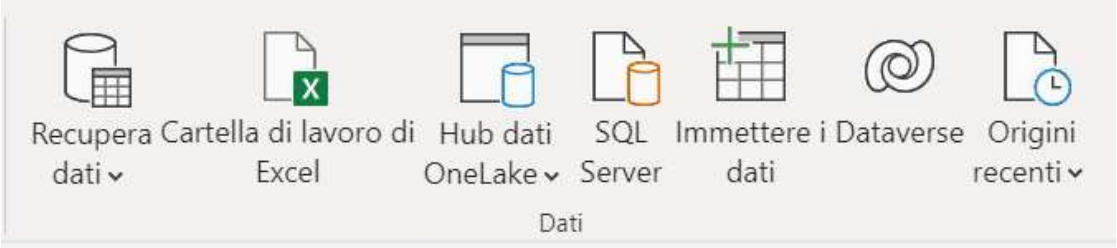

Figura 3.8: Barra superiore di Power BI dedicata all'acquisizione di dati da varie sorgenti

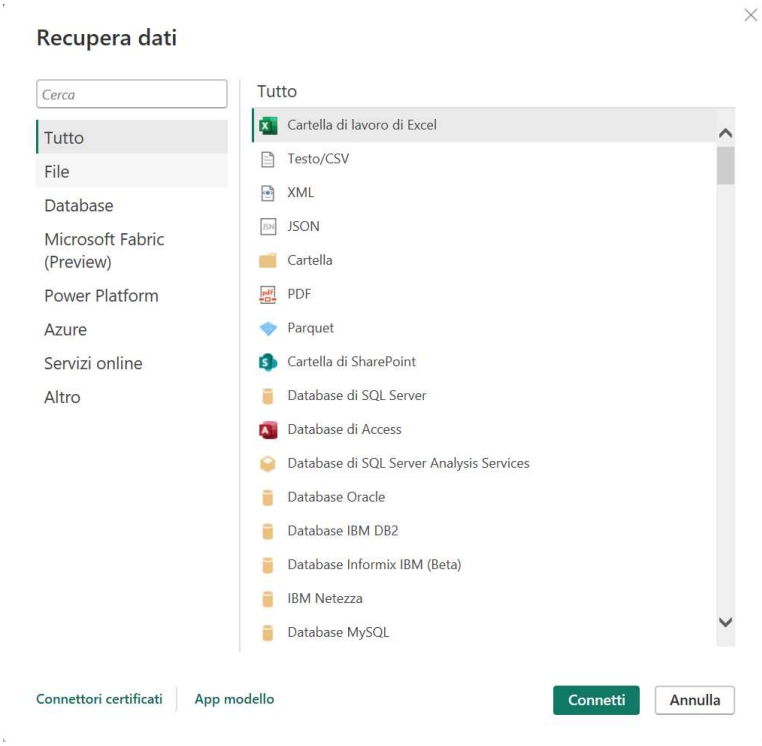

Figura 3.9: Finestra *Recupera dati* in Power BI Desktop

3.4.2 Transform

Una volta completato il processo di estrazione, procediamo con la fase di trasformazione. Questo stadio del processo consiste principalmente nel pulire i dati e strutturarli in una forma che ci permetta di creare report significativi. Innanzitutto, ci assicuriamo che tutti i dati degli attributi, ossia le colonne, siano presentati nel formato corretto; nel nostro caso Power BI ha effettuato una conversione dei tipi in automatico; in particolare a tutti gli attributi è stato assegnato il tipo String poiché sul foglio di lavoro in Excel i dati erano principalmente salvati come String. Quindi è stato necessario effettuare una conversione di tipo per tutti gli attributi; ad esempio:

- il tipo del campo "dt" è stato convertito nel tipo "datetime";
- il tipo dell'attributo "annulled" è stato convertito nel tipo "datetime";
- il tipo del campo "PREZZOL20" è stato convertito nel tipo "int";
- il tipo dell'attributo "cliente" è stato convertito nel tipo "int";
- il tipo del campo "brand" è stato convertito nel tipo "text".

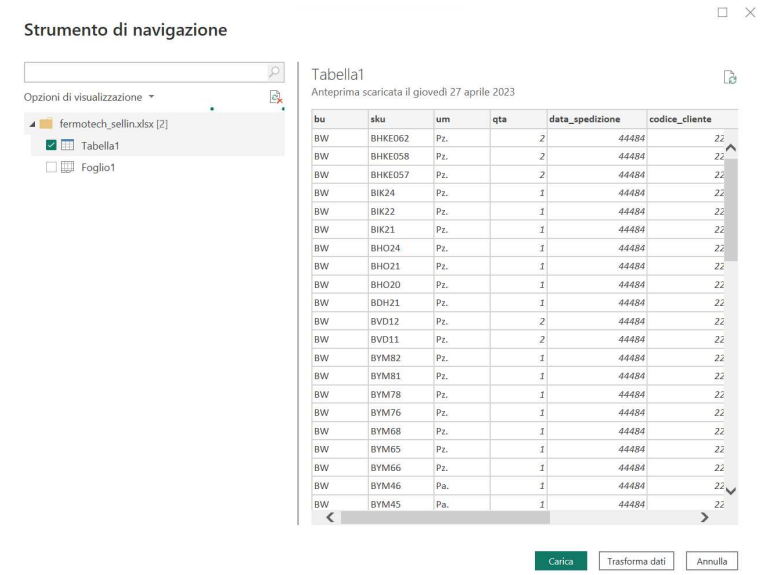

Figura 3.10: Finestra *strumento di navigazione* in Power BI Desktop

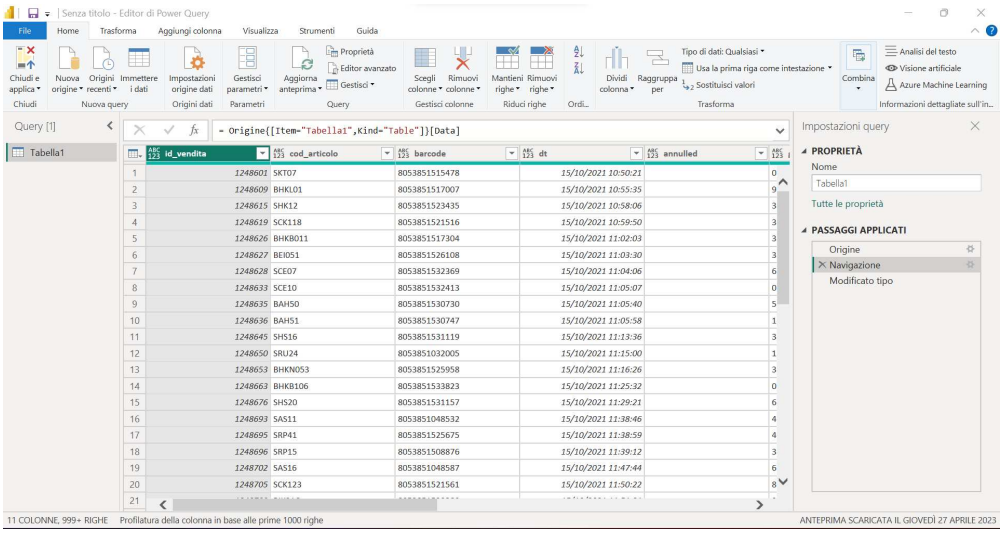

Figura 3.11: Schermata della tabella iniziale in Power Query

Successivamente, verifichiamo l'accuratezza delle intestazioni; sotto questo aspetto, non si riscontra nessun problema. Procedendo con l'ispezione, abbiamo osservato che alcuni attributi presentano delle incongruenze, come illustrato in Figura 3.12. Queste irregolarità sono evidenziate anche dalla presenza di colorazioni rosse sulla barra inferiore, che solitamente è completamente verde. In Power BI l'operazione di rimozione degli errori è presente di default nella barra delle applicazioni. Quindi, per ogni colonna contenente errori, è sufficiente richiamare l'apposita funzione.

3.4.3 Load

Dopo aver effettuato le modifiche necessarie attraverso il processo di trasformazione, arriviamo alla fase di caricamento dei dati. In questa specifica situazione, i dati sono spostati dall'editor Power Query direttamente all'interno di Power BI, dove saranno quindi accessibili per l'elaborazione di report. Per completare questo procedimento, è sufficiente cliccare sull'opzione "*Chiudi e Applica*", come illustrato nella Figura 3.13.

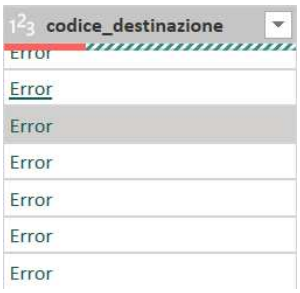

Figura 3.12: Schermata di una colonna contenente degli errori

Figura 3.13: Opzione "*Chiudi e applica*" di PowerQuery

CAPITOLO 4

Analisi effettuate e risultati derivati

In questo capitolo, dopo un breve accenno alle varie categorie di Data Analytics, ci soffermeremo in maniera particolare sull'analisi descrittiva. Analizzeremo dettagliatamente le procedure di analisi dei dati eseguite sui dataset aziendali forniti da Bros Manifatture. Il nostro obiettivo principale è descrivere le analisi effettuate esaminando i report realizzati in Power BI, dopo aver raccolto ed elaborato i dati a nostra disposizione. Le nostre analisi saranno, in particolar modo, indirizzate verso lo studio del cliente e del consumatore finale.

4.1 Introduzione ai tipi di analisi

Come delineato nei capitoli precedenti, lo scopo primario dell'analisi dei dati è quello di produrre informazioni utili per guidare le decisioni aziendali. È importante, però, precisare che ci sono varie categorie di analisi dei dati, come illustrato nella Figura 4.1. Naturalmente, attraverso un'analisi più dettagliata e complessa otterremo informazioni più significative per l'organizzazione.

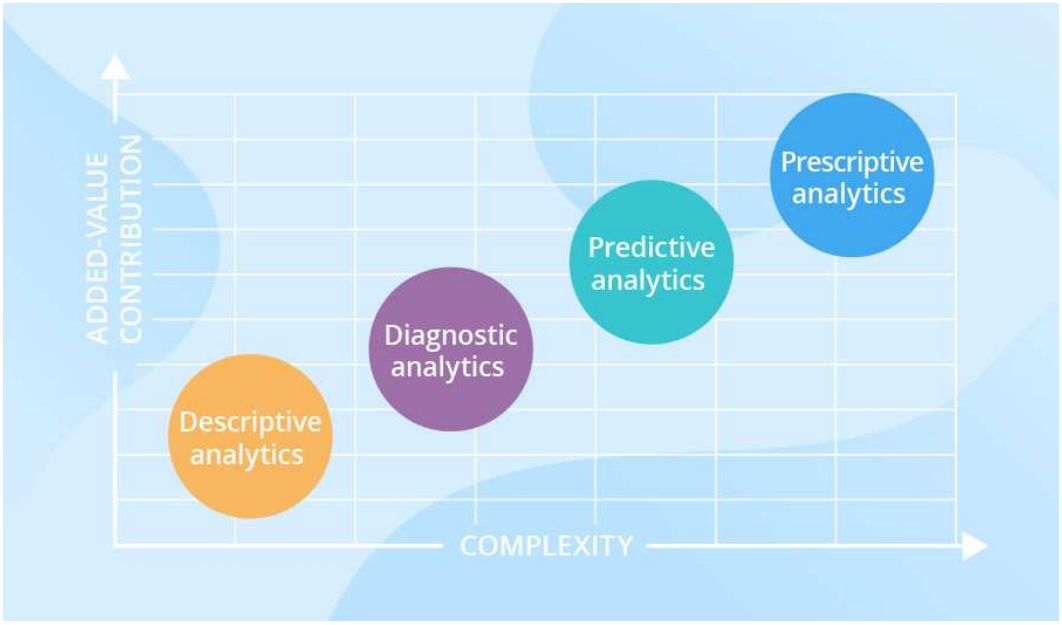

Figura 4.1: Le 4 tipologie della Data Analytics

Si distinguono quattro principali tipologie dell'analisi dei dati, classificate in base ai risultati che generano:

- *Analisi descrittiva*: l'obiettivo dell'analisi descrittiva è fornire risposte a interrogativi riguardanti eventi avvenuti in passato.
- *Analisi diagnostica*: l'obiettivo dell'analisi diagnostica è identificare le cause sottostanti a un evento che si è verificato in precedenza.
- *Analisi predittiva*: l'analisi predittiva si pone l'obiettivo di anticipare gli eventi che potrebbero verificarsi in futuro.
- *Analisi prescrittiva*: l'analisi prescrittiva, che si appoggia sui risultati dell'analisi predittiva, ha l'obiettivo di suggerire le azioni da adottare in futuro.

Nella prossima sezione illustreremo in modo più approfondito l'analisi descrittiva.

4.1.1 Analisi descrittiva

L'analisi descrittiva fornisce risposte relative a eventi che si sono verificati in passato, permettendo, così, una comprensione di questi attraverso l'elaborazione dei dati raccolti. Questo significa che essa riesce a trasformare dati grezzi in informazioni utili, fornendo un quadro d'insieme di ciò che è avvenuto. Questo tipo di analisi utilizza strumenti di Data Visualization per mostrare visivamente in modo intuitivo la grande mole di dati storici. In questo contesto, si utilizzano report e dashboard personalizzati che offrono una visione immediata e significativa dei dati, evidenziando schemi, correlazioni e trend nascosti. Di solito, la preferenza viene data ad una rappresentazione visiva e immediata dei dati piuttosto che a tabelle con un gran numero di righe. Quindi, le informazioni utili possono essere dedotte osservando le tendenze grafiche dei dati. Ad esempio, si potrebbe osservare che nel mese di gennaio le vendite tendono ad essere più basse, mentre dicembre sembra essere il periodo più redditizio.

Ecco alcuni esempi di domande tipiche a cui può rispondere l'analisi descrittiva:

- Qual è il trend delle vendite nel corso del tempo?
- Quali prodotti o servizi sono più venduti?
- Quali sono i periodi di punta per le vendite?
- Quale canale di vendita è più efficace?

Essenzialmente, l'analisi descrittiva può aiutare a rispondere a qualsiasi domanda che riguardi il "chi", "cosa", "dove" e "quando" dei dati aziendali raccolti.

Nelle sezioni successive, analizzeremo in maniera dettagliata le procedure di analisi dei dati eseguite sui dataset aziendali forniti da Bros Manifatture. L'obiettivo principale è quello di descrivere le analisi effettuate e commentare i report realizzati dopo aver raccolto ed elaborato i dati a nostra disposizione.

4.2 Analisi effettuate sul cliente

In questa specifica sezione, eseguiremo un'indagine dettagliata sui report di analisi e visualizzazione riguardanti i clienti. Saranno oggetto di particolare interesse vari fattori statistici quali le quantità dei vari prodotti spediti e consegnati ai vari clienti e i pattern di acquisto in relazione alle diverse stagioni dell'anno.

4.2.1 Report stagionalità

Nella Figura 4.2 abbiamo un istogramma a colonne in pila; sull'asse delle ascisse sono riportati i mesi da gennaio 2022 a marzo 2023 mentre sull'asse delle ordinate troviamo la quantità dei prodotti spediti ai vari clienti.

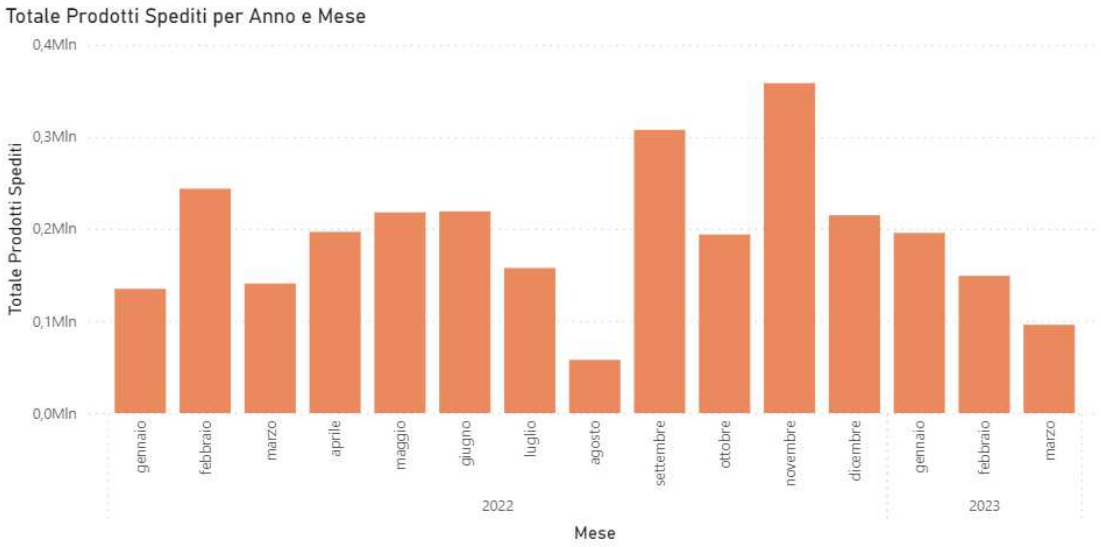

Figura 4.2: Istrogramma a pila dei prodotti spediti in base al periodo dell'anno

Dall'osservazione di questo report è possibile dedurre le seguenti informazioni:

- *Periodicità stagionale*: prima di tutto, da quanto osservato, sembra esserci un certo grado di stagionalità nei dati. Febbraio, settembre e dicembre sono i mesi con i volumi di spedizione più elevati. Questo suggerisce un possibile picco delle vendite durante il periodo di San Valentino, l'autunno e il periodo natalizio. Al contrario, i periodi di volume più basso sono agosto, marzo e gennaio. Si può ipotizzare una fase di calo delle vendite dopo i periodi festivi e durante l'estate.
- *Trend annuale*: rispetto all'andamento generale dell'anno, notiamo una diminuzione dei prodotti spediti nei mesi di agosto e di marzo dell'anno seguente, il che potrebbe indicare l'esistenza di un trend di calo degli acquisti in quei particolari mesi.
- *Variazioni mensili*: è possibile indagare, inoltre, sulle variazioni mensili nel volume di prodotti spediti. Ad esempio, sebbene gennaio abbia un volume di spedizioni piuttosto basso rispetto ad altri mesi, è comunque notevole se confrontato con agosto o marzo. Le altre colonne sono tutte di altezza media e molto simili tra loro; esse indicano, quindi, che le spedizioni sono relativamente stabili durante la maggior parte dell'anno.

4.2.2 Report quantità spedite

Nella Figura 4.3 analizziamo un ulteriore istogramma a colonne in pila. Esso fornisce un'immagine dettagliata della distribuzione dei clienti dell'azienda in base al volume degli ordini. Da questo grafico è evidente che l'azienda Bros Manifatture ha un'ampia gamma di clienti in termini di volume di prodotti ordinati.

Gli ordini spediti ai clienti segnano, infatti, una significativa diversità; in particolare è presente un cliente principale (18500) che ha ricevuto un numero notevolmente più alto di prodotti (424mila) rispetto a tutti gli altri clienti. Gli ordini ricevuti dai restanti clienti

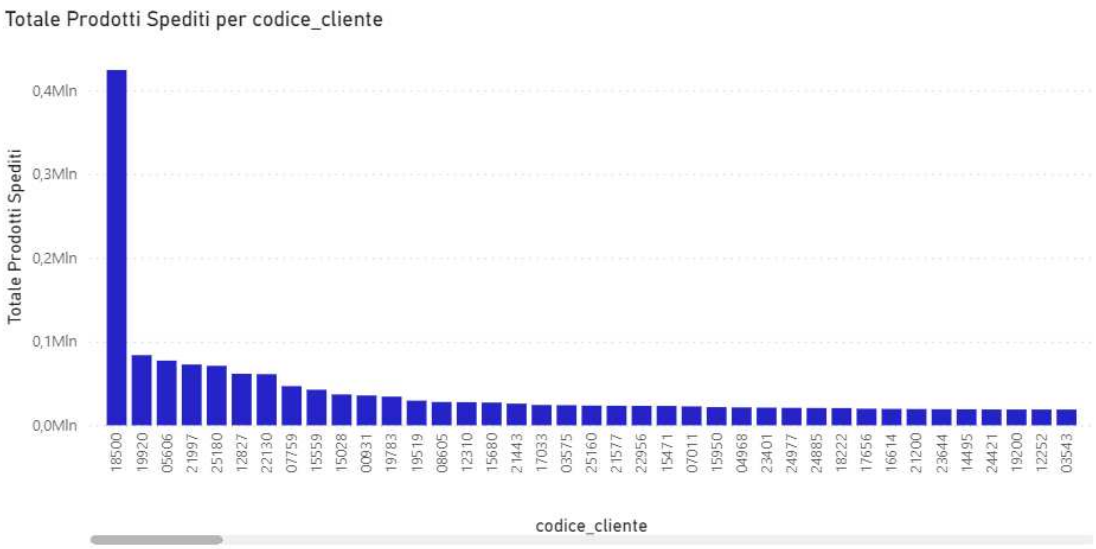

Figura 4.3: Estratto selezionato dell'istogramma a pila sui prodotti spediti in base al cliente

sembrano avere una distribuzione più omogenea, con una tendenza verso clienti che ordinano una quantità di prodotti nel range dei 10k-15k.

Tra i clienti di punta abbiamo cinque di loro che hanno ricevuto ordini tra 84k e 60k prodotti, seguiti da altri sei tra 47k e 29k. Questo suggerisce che c'è un piccolo numero di clienti che piazzano ordini di quantità elevate.

La maggioranza dei clienti, circa 70, si trovano nella fascia 10k-15k. Questi effettuano numerosi ordini, ma in quantità minori rispetto alle punte; possiamo quindi definirli "clienti frequenti".

Allo stesso tempo, c'è un numero significativo di clienti che effettuano ordini più piccoli, tra 600 e 8 prodotti, che potrebbero rappresentare clienti occasionali, o piccoli rivenditori.

In generale, la distribuzione dei clienti per prodotti spediti mostra una tendenza decrescente, con il numero di clienti che aumenta man mano che il numero di prodotti spediti per cliente diminuisce. Questo è un fenomeno comune in molti settori di vendita al dettaglio e indica che l'azienda ha un ampio numero di clienti che effettuano ordini minori, equilibrato da pochi clienti di punta che effettuano ordini di grandi quantità.

4.2.3 Report in base al brand

Nella Figura 4.4 viene mostrato un grafico a torta che fornisce un'istantanea della ripartizione percentuale dei prodotti spediti dei diversi brand dell'azienda Bros. Anche se vi sono sei brand in questione, tre di essi rappresentano la grande maggioranza delle spedizioni.

Il brand "BW" è chiaramente quello di punta dell'azienda, rappresentando circa il 59,39% del totale dei prodotti spediti. Questo è probabilmente dovuto a una combinazione di fattori come la qualità del prodotto, la popolarità del brand e una possibile varietà di offerte.

Con il 24,63% dei prodotti spediti, "SA" è chiaramente un brand molto importante per l'azienda, anche se non raggiunge la popolarità di BW. Potrebbe essere un marchio più specializzato, di seconda scelta per i rivenditori, o semplicemente avere un seguito minore rispetto a BW.

Rappresentando il 12,28%, DH potrebbe essere un brand di nicchia che mira a un segmento specifico di clienti. È un contributo abbastanza considerevole alle vendite totali, anche se non è dominante.

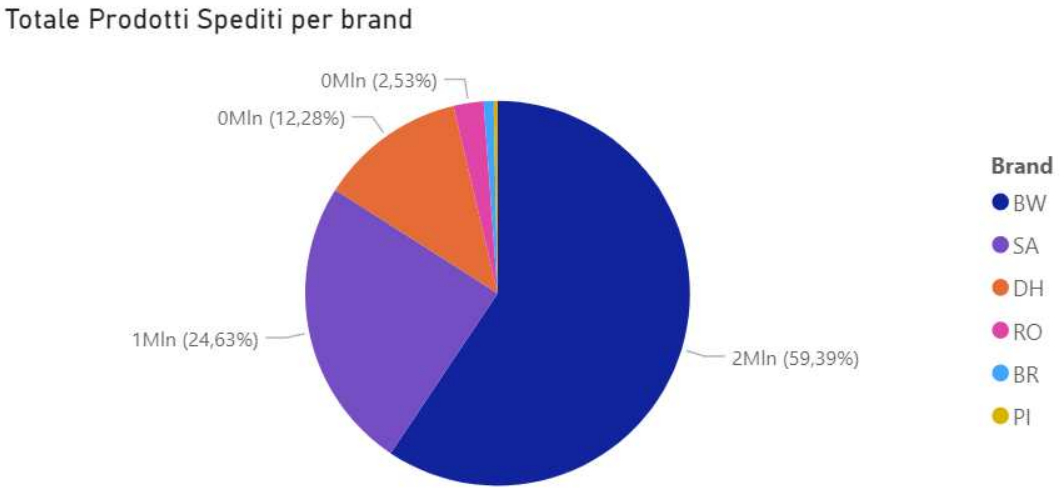

Figura 4.4: Grafico a torta dei prodotti spediti in base al brand

Passando ai brand restanti, essi rappresentano insieme solo l'8,69% delle spedizioni. Il brand RO rappresenta il 2,53%, seguito da BR con lo 0,83% e PI con lo 0,33%. Ciò potrebbe indicare che sono brand meno noti o specializzati in prodotti di nicchia con una domanda limitata. Sono piccoli contributori alle vendite totali ma potrebbero avere un potenziale di crescita.

Nel complesso, l'azienda Bros sembra essere altamente dipendente dal brand BW e in misura minore da SA e DH. I brand RO, BR e PI contribuiscono minimamente alle vendite totali. Questa è una possibile area di preoccupazione in termini di diversificazione del portafoglio del brand e riduzione del rischio. L'azienda Bros potrebbe guardare alle strategie per incrementare la popolarità e la vendita dei brand di minore contribuzione al fine di ottenere una distribuzione più equilibrata delle vendite tra i brand.

4.3 Analisi effettuate sul consumatore finale

In questa sezione, effettueremo un'indagine dettagliata sui report di analisi e visualizzazione riguardanti le vendite al consumatore finale. Analizzeremo, in modo particolare, diversi fattori statistici, come il volume degli articoli acquistati e le tendenze di acquisto in relazione alle varie stagioni dell'anno.

4.3.1 Report sulle vendite

Iniziamo la nostra analisi con un grafico a nastro, in Figura 4.5, il quale ci svela gli andamenti stagionali nelle vendite al consumatore finale in rapporto alle vendite annullate. Sull'asse sono riportati i mesi da ottobre 2021 a marzo 2023. Abbiamo due nastri, ovvero il nastro arancione, che rappresenta le vendite totali al consumatore finale, e il nastro blu, che rappresenta le vendite annullate o gli articoli restituiti.

Analizziamo i due nastri nel dettaglio:

- *Nastro arancione*: le vendite al consumatore finale mostrano uno schema stagionale significativo, con picchi evidenti nei mesi di dicembre e una diminuzione nei mesi di gennaio. Il picco di dicembre può rappresentare l'effetto delle vendite natalizie sulle vendite al dettaglio, con un aumento delle spese dei consumatori. La diminuzione di

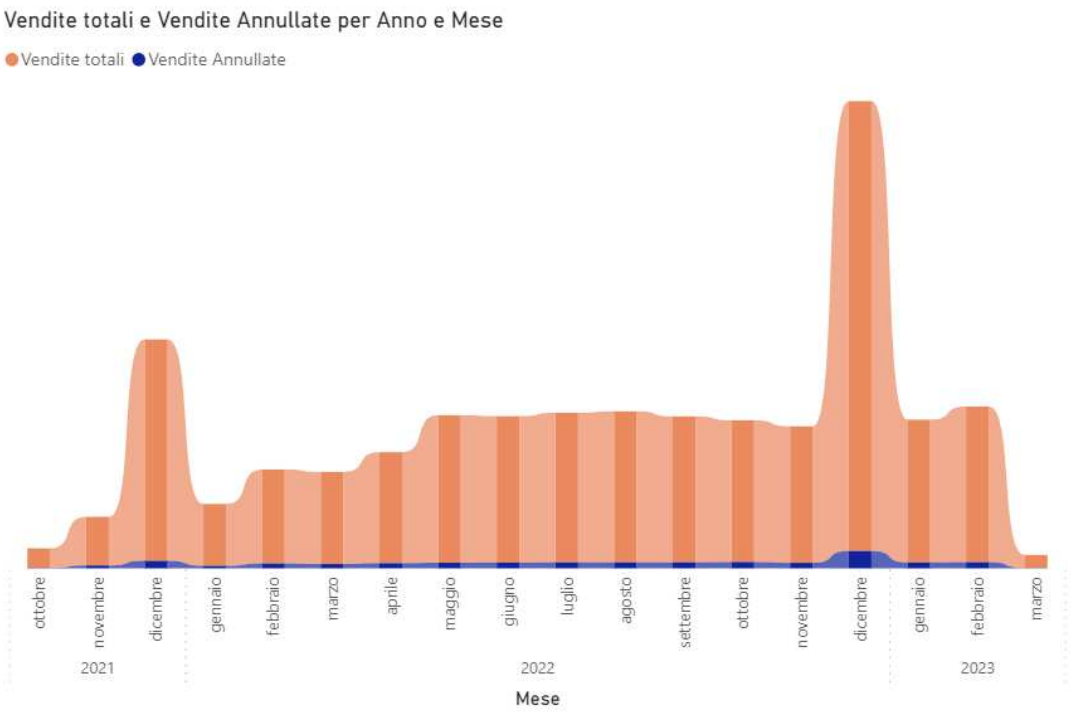

Figura 4.5: Grafico a nastro delle vendite totali e delle vendite annullate

gennaio potrebbe essere attribuita a una riduzione delle spese dopo il periodo festivo. A partire da maggio 2022 fino a novembre dello stesso anno, si nota una sostanziale uniformità dei dati di vendita, sebbene siano più bassi rispetto al picco di dicembre 2022. Il periodo da febbraio a maggio 2022 mostra un aumento leggermente più basso rispetto agli altri periodi, suggerendo che potrebbe esserci un periodo di vendite più lente in questi mesi.

- *Nastro blu*: il nastro blu, che rappresenta le vendite annullate o restituite, è generalmente molto più sottile rispetto al nastro arancione, indicando che solo una piccola percentuale delle vendite totali viene annullata o restituita. Tuttavia, questo nastro mostra anche dei picchi in dicembre 2021 e dicembre 2022, suggerendo che le restituzioni o le cancellazioni possono aumentare in concomitanza con il picco delle vendite. Questo può essere dovuto all'aumento degli acquisti regalo che vengono, poi, restituiti o scambiati dopo le festività.

Questo tipo di analisi aiuta a identificare le tendenze di vendita in relazione al periodo dell'anno, con un picco durante le festività natalizie. Inoltre, segnala che la maggior parte delle vendite rimane non restituita o annullata. Le tendenze nelle vendite annullate sembrano seguire il trend generale delle vendite al dettaglio.

4.3.2 Report sui prodotti

Nella Figura 4.6 vi è una mappa ad albero che mette in evidenza le vendite totali in correlazione con il codice del prodotto. Prendendo in considerazione gli articoli più venduti, si può notare che:

- Le linee di prodotti BEI e BHK dominano le vendite totali ai clienti. Ciò indica che i prodotti che appartengono a queste collezioni sono estremamente popolari tra i

clienti, probabilmente grazie al loro design classico ed elegante per BEI, e i simboli che raccontano storie di fortuna, amicizia e amore per BHK.

- La linea di prodotti SHA ha una partecipazione più piccola nel complesso; il fatto che compaia tra le celle più grandi suggerisce che ha, tuttavia, un solido seguito di clienti. I suoi disegni marini e il suo design casual possono richiamare un segmento di clienti in cerca di un'estetica diversa.
- La popolarità di BEI e BHK potrebbe anche riflettere una preferenza generale dei clienti per i gioielli che combinano sia l'eleganza che un senso di personalità o narrazione data dai vari simboli.

Figura 4.6: Estratto della mappa ad albero dei prodotti venduti

4.3.3 Report finale

Concludiamo la nostra campagna di Data Analytics con la visualizzazione di un albero di scomposizione (Figura 4.7).

Come sappiamo, quest'ultimo è uno strumento di Data Visualization; esso scompone il nodo principale in sotto-categorie al fine di fornire una comprensione più dettagliata dei dati.

L'albero in figura può essere descritto come segue:

- *Nodo primario*: questo è il primo e più tassativo livello del nostro albero di scomposizione. Esso rappresenta la quantità totale di acquisti effettuati dal consumatore finale.

- *Nodo secondario*: questo nodo scompone le vendite totali in base alla data. Le vendite possono essere scomposte annualmente, mensilmente, settimanalmente o anche giornalmente. Permette la visualizzazione dell'andamento delle vendite nel tempo.
- *Nodo terziario*: sotto il nodo data, abbiamo il nodo brand. Quest'ultimo divide ulteriormente le vendite a seconda del brand. Esso ci mostra qual è il brand più popolare in un determinato periodo di tempo.
- *Nodo quaternario*: questo è il livello più dettagliato del nostro albero di scomposizione. Suddivide le vendite per brand in base al codice dell'articolo, permettendo, quindi, la visualizzazione delle vendite di prodotti specifici di ciascun brand. Questo offre una visione estremamente dettagliata delle vendite, mostrando quali prodotti sono più popolari tra i consumatori.

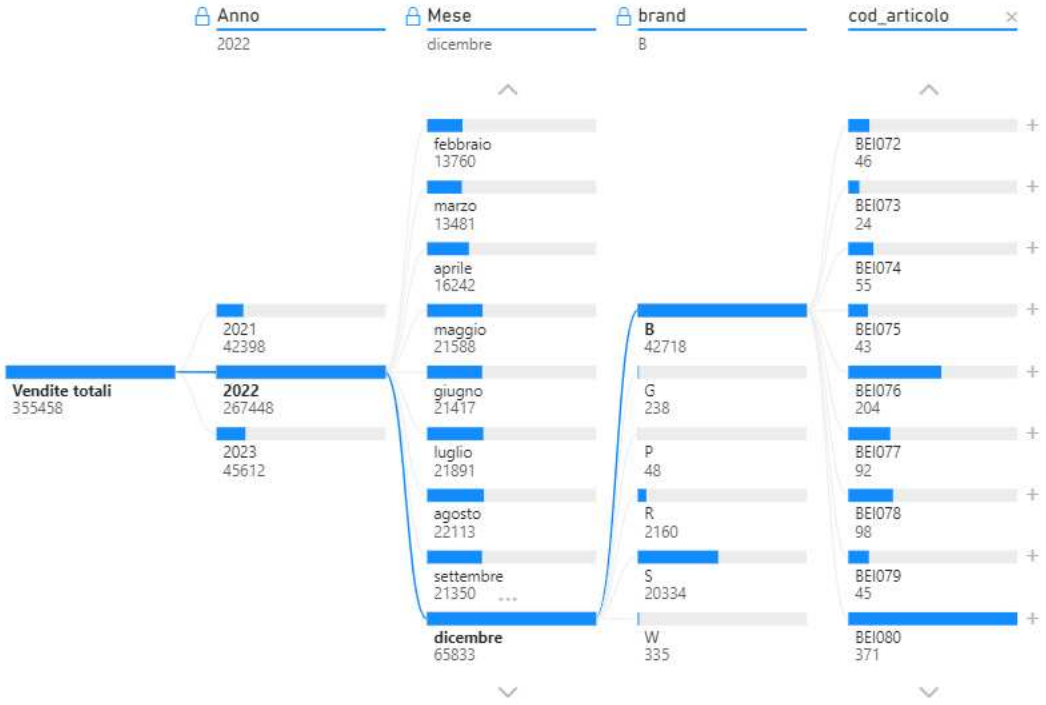

Figura 4.7: Albero di scomposizione sulle vendite totali

I risultati dell'analisi di questo albero ci danno la conferma che quanto abbiamo descritto nella sezione riguardante le vendite alle attività commerciali si riflette sulle vendite al consumatore finale. Infatti, possiamo notare che le vendite raggiungono il loro picco massimo sempre nel mese di dicembre mentre il brand più popolare risulta essere sempre "B". Sottolineiamo, inoltre, che, osservando il nodo riguardante il codice dell'articolo, risulta confermata l'analisi sui prodotti in quanto gli articoli più venduti al cliente finale corrispondono con le analisi precedenti.

CAPITOLO 5

Discussione in merito al lavoro svolto

Nel presente capitolo, cercheremo di approfondire la tematica della Business Intelligence, analizzando il suo impatto sul panorama aziendale odierno, sempre più incentrato sull'analisi dei dati. L'attenzione si sposterà poi sui risultati ottenuti attraverso l'analisi del percorso di Big Data Analytics intrapreso, enfatizzando sia gli aspetti positivi che quelli negativi. In questo modo, intendiamo offrire una visione completa del lavoro svolto.

5.1 Business Intelligence: impatto ed esigenze nel panorama aziendale

In questo capitolo, ci proponiamo di discutere i risultati ottenuti e di valutare gli aspetti positivi della nostra campagna di Data Analytics nonchè gli ostacoli incontrati. L'analisi dei dati e la comprensione delle tendenze nascoste in essi è un elemento critico nel mondo degli affari di oggi; il nostro lavoro ha sottolineato l'importanza e il valore della Business Intelligence e della Data Analytics in questo contesto.

Nel panorama aziendale attuale, altamente digitalizzato, l'accesso a dati solidi e affidabili rappresenta una risorsa chiave per le imprese tese ad ottenere un vantaggio competitivo. A tal proposito, l'attività o processo di acquisizione, formattazione, manipolazione, analisi e presentazione dei dati aziendali si traduce in quella che viene conosciuta comunemente come Business Intelligence (BI).

La Business Intelligence comprende un insieme di metodologie, applicazioni e tecnologie che consentono alle aziende di raccogliere, conservare e analizzare dati per aiutare gli utenti aziendali a prendere decisioni informate.

All'interno della Business Intelligence si annidano concetti come data warehousing, data mining, reporting e analisi di business. Al centro della BI vi è un elemento chiave: trasformare i dati grezzi in informazioni significative. Tuttavia, ricavare informazioni preziose dai dati non è un compito facile. La BI, quindi, serve come veicolo per aiutare le aziende a navigare in un mare di dati complessi.

In passato, i metodi tradizionali di presa delle decisioni nel mondo degli affari si basavano spesso su esperienza ed intuito. Oggi, l'ascesa della Business Intelligence ha determinato un cambiamento radicale. La dipendenza dai dati ha portato all'affermazione del concetto di "data-driven decision-making" (ovvero prendere decisioni basandosi sui dati), dal momento che l'utilizzo di informazioni quantitative è sempre più importante rispetto all'esperienza e all'intuizione personale di chi prende decisioni.

Le aziende si stanno evolvendo verso un modello sempre più "data centric", riconoscendo i dati come una risorsa strategica fondamentale per migliorare le proprie prestazioni. Questo flusso di evoluzione aziendale conferma l'importanza dell'affermazione di Lord William Kelvin, il grande fisico ed ingegnere britannico, il quale sosteneva che "se non si può misurare qualcosa, non si può migliorarla". Infatti, nel contesto di un'impresa "data centric", le decisioni strategiche ed operative sono guidate dalla misurazione e dall'analisi dei dati, permettendo così di identificare le aree di miglioramento e di intervenire in modo mirato ed efficace.

5.2 Aspetti positivi e negativi del lavoro svolto

Nel presente paragrafo si intendono esplorare i risultati ottenuti mediante il percorso di analisi condotto sulla campagna di Big Data Analytics realizzata per studiare le vendite di un importante produttore di gioielli. Attraverso una serie di strumenti digitali ed una metodologia d'approccio con la Business Intelligence (BI), si è cercato di capire meglio le dinamiche di mercato del settore in esame con l'obiettivo di migliorare le performance di vendita e di business.

Il punto di partenza è stato un set di dati relativamente pulito e di facile elaborazione; questo ha permesso di adoperare una soluzione software come Power BI senza l'impiego del linguaggio di programmazione Python per le operazioni preliminari di pulizia e organizzazione dei dati. Questo aspetto ha contribuito notevolmente nell'abbreviare le tempistiche di analisi, consentendo un più agevole raggiungimento degli obiettivi. Inoltre, Power BI ha permesso di eseguire analisi dei dati efficienti e di elaborare report di alto livello che forniscono una visualizzazione chiara e sintetica dei risultati.

Tuttavia, sebbene il nostro lavoro presenti diversi punti di forza, ci sono alcune aree che avrebbero potuto beneficiare di un'attenzione maggiore. Uno degli aspetti limitanti del nostro studio è stata l'assenza di dati geografici nei dataset forniti. Ciò ha ridotto la possibilità di eseguire analisi spaziali, che avrebbero potuto rivelare tendenze geografiche nei dati di vendita. Inoltre, è importante sottolineare come una maggiore comunicazione con l'azienda ci avrebbe permesso di integrare meglio le analisi eseguite e di ottenere risultati più ricchi e significativi.

Nonostante questi limiti, riteniamo che il lavoro svolto abbia comunque fornito un quadro analitico solido e valido. Le analisi descrittive realizzate, indirizzate a studiare il cliente e il consumatore finale, hanno offerto una visione chiara dei pattern di vendita, contribuendo a isolare le tendenze principali da prendere in considerazione per le strategie di marketing future. È fondamentale sottolineare come le analisi effettuate, pur non essendo esenti da limitazioni, siano un passo importante nel creare valore da grandi set di dati e nel guidare decisioni informate basate su di essi.

In conclusione, ciò che emerge dal nostro lavoro è l'importanza insostituibile della Business Intelligence e della Data Analytics nel contesto commerciale moderno. Questi strumenti, ben utilizzati, hanno il potenziale di fornire intuizioni valide, guidare verso decisioni efficaci e sostenere la crescita e il successo delle imprese.

Conclusioni e uno sguardo al futuro

Il cammino intrapreso all'interno di questa tesi, focalizzato sulla conduzione di una campagna di Data Science applicata alla vendita di prodotti di un'importante azienda di gioielli, ha permesso di presentare e approfondire complesse tematiche relative alla gestione e analisi dei Big Data, esemplificando le potenzialità offerte da tale approccio tramite l'uso di Power BI.

Siamo partiti offrendo una completa panoramica sulla Data Analytics, delineando il quadro generale dei Big Data, passando dalla teoria delle 3V di Laney fino a quella delle 5V, e attraversando l'intero ciclo di vita della Big Data Analytics. In seguito, abbiamo esplorato Power BI, evidenziando le sue peculiarità, e abbiamo esposto il ruolo fondamentale ricoperto dai vari tipi di dati nell'analisi delle informazioni. Abbiamo approfondito il processo di ETL (Extract, Transform, Load), essenziale nel campo del data warehousing. Infine, ci siamo concentrati sull'analisi descrittiva dei dati, prendendo in considerazione i dati di vendita dell'azienda marchigiana Bros Manifatture.

Nonostante mancassero alcuni elementi fondamentali, come i dati geografici, siamo riusciti comunque a realizzare vari tipi di report tra cui report sulla stagionalità delle vendite, sulle vendite in base al brand e sulle vendite effettuate rispetto a quelle annullate, fornendo, così, un quadro sul comportamento del consumatore.

Malgrado le limitazioni riscontrate, l'approccio adottato e le tecniche utilizzate hanno fornito una sostanziale comprensione dei pattern di vendita, illuminando importanti settori dell'azienda che meritano attenzione e possibili riscontri.

Rivolgendo uno sguardo al futuro, è sicuramente possibile un ampliamento di questa campagna di Data Analytics. Si può pensare di implementare analisi predittive, attraverso l'uso di algoritmi di Intelligenza Artificiale e machine learning, per previsioni sulle vendite o per scoprire nuovi orientamenti del consumatore. Non si escludono possibili analisi prescrittive che permettano all'azienda di prendere decisioni basate su scenari futuri simulati.

Un collegamento più stretto con l'azienda potrebbe permettere un aggiornamento costante dei dataset e l'inserimento di ulteriori informazioni preziose, come, appunto, i dati geografici, posizionando l'analisi su un livello superiore.

Questa tesi può essere considerata un punto di partenza per progetti futuri o un modello di base per altre aziende che desiderano rivolgere lo sguardo verso il mondo della Data Analytics. A fronte di un mondo sempre più "data centric" si auspica una maggiore diffusione di tale approccio, rendendo le organizzazioni sempre più consapevoli del valore dei dati, non solo come numeri, ma come fonti inesauribili di conoscenza e di nuove opportunità.

Bibliografia

DE MAURO, A., GRECO, M. e GRIMALDI, M. (2016), «A formal definition of Big Data based on its essential features», *Library Review*, vol. 65 (3), p. 122–135, URL https://doi.org/10.1108/LR-06-2015-0061. (Cited at page 3)

DI NUZZO, M. (2021), *Data Science e Machine Learning: Dai Dati alla Conoscenza*, Michele di Nuzzo, URL https://books.google.it/books?id=8KVCEAAAQBAJ.

FERRARI, A. e RUSSO, M. (2016), *Introducing Microsoft Power BI*, Pearson Education, URL https://books.google.it/books?id=U1qsDAAAQBAJ.

LANEY, D. (2001), «3D Data Management: Controlling Data Volume, Velocity, and Variety», Rap. tecn., META Group.

RAD, R. e ETAATI, L. (2021), *Mastering Power Query in Power BI and Excel: Learning real-world Power Query and M Techniques for a better data analysis*, RADACAD Systems Limited, URL https://books.google.it/books?id=iKs_EAAAQBAJ.

REINSEL, D., GANTZ, J. e RYDNING, J. (2017), «Data Age 2025: The Evolution of Data to Life-Critical. Don't Focus on Big Data; Focus on the Data That's Big», Rap. tecn., International Data Corporation (IDC).

SALVAGGIO, A. (2023), *Business Intelligence con Microsoft Power BI: Guida completa per l'analisi e la visualizzazione dei dati*, Edizioni LSWR, URL https://books.google.it/books?id=j7jXzwEACAAJ.

SHARDA, R., DELEN, D. e TURBAN, E. (2020), *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*, Pearson, URL https://books.google.it/books?id=teRIzQEACAAJ.

SPECTOR, A., NORVIG, P., WIGGINS, C. e WING, J. (2022), *Data Science in Context: Foundations, Challenges, Opportunities*, Cambridge University Press.

VERHOEF, P. C., KOOGE, E. e WALK, N. (2016), *Creating Value with Big Data Analytics: Making Smarter Marketing Decisions*, Taylor & Francis, URL https://books.google.it/books?id=w7RYCwAAQBAJ.

Siti web consultati

- Microsoft Power BI – `https://docs.microsoft.com/en-us/power-bi/`
- Bros Manifatture – `www.brosmanifatture.it`
- Gartner – `www.gartner.com`
- SQLBI – `www.sqlbi.com`
- Brosway – `www.brosway.com`

Ringraziamenti

Vorrei esprimere la mia sincera gratitudine a coloro che hanno reso questo percorso universitario un'esperienza significativa e gratificante.

Prima di tutto, vorrei ringraziare il mio relatore, il Professore Domenico Ursino, per la sua dedizione, la sua pazienza e per avermi guidato lungo questo cammino. Le sue preziose indicazioni e i suoi consigli hanno reso questo lavoro molto più agevole.

Un ringraziamento speciale anche al Professore Francesco Cauteruccio, il mio correlatore, per i suoi preziosi consigli e per avermi supportato nell'elaborazione di questo lavoro. La sua assistenza ha arricchito il mio percorso di studi.

Desidero esprimere la mia immensa gratitudine ai miei genitori e a mio fratello Aldo, per il loro affetto infinito e per avermi sempre sostenuto in ogni decisione. Il vostro amore e sostegno sono stati fondamentali in questo lungo percorso.

Ringrazio i miei zii, in particolare la carissima zia Patrizia, per la sua presenza costante e per avermi incoraggiato nel conseguimento di questo traguardo.

Vorrei ringraziare i miei amici, in particolare Flavia, Hermes e Regy. Il vostro supporto, le risate, i consigli e la presenza in ogni momento mi hanno aiutato a superare i giorni più difficili.

Un ringraziamento speciale al mio fidanzato Lorenzo. Nonostante la lontananza, hai sempre trovato il modo di essere presente, di incoraggiarmi e di sostenere i miei sforzi. La tua forza e il tuo amore mi hanno dato l'energia necessaria per portare a termine questo percorso.

Infine, ma non meno importante, desidero esprimere il mio più profondo affetto e riconoscimento ai miei nonni. La vostra saggezza, il vostro amore e il vostro sostegno sono stati un faro di luce lungo questo percorso. Il vostro esempio mi ha ispirato ad essere una persona migliore e a non arrendermi mai di fronte alle difficoltà. Grazie amatissima nonnina che anche da lassù sei orgogliosa di me.

A tutti voi, il mio più sincero ringraziamento.